STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 176372

TO: Konstantina Katcheves **Location: REM-2A60/2C70** **Art Unit: 1636** **Friday, January 20, 2006** **Case Serial Number: 10/511327**	**From: Barb O'Bryen** **Location: Biotech-Chem Library** **Remsen 1a69** **Phone: 571-272-2518** **barbara.obryen@uspto.gov**

Search Notes

RUSH

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 278.233 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-1
Perfect score: 54
Sequence: 1 gatctgattgattgaacgaa.....ttcacgataaatgcgaaaac 54

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42.8	79.3	1319	11	AY781404	AY781404 Synthetic
c 2	42.8	79.3	1319	11	AY781404	AY781404 Synthetic
3	42.8	79.3	1351	11	AY781405	AY781405 Synthetic
c 4	42.8	79.3	1351	11	AY781405	AY781405 Synthetic
5	42.8	79.3	1524	11	AY781401	AY781401 Synthetic
c 6	42.8	79.3	1524	11	AY781401	AY781401 Synthetic
7	42.8	79.3	1808	11	AY781408	AY781408 Synthetic
c 8	42.8	79.3	1808	11	AY781408	AY781408 Synthetic
9	42.8	79.3	1937	11	AY781403	AY781403 Synthetic
c 10	42.8	79.3	1937	11	AY781403	AY781403 Synthetic
11	42.8	79.3	2037	11	AY781407	AY781407 Synthetic
c 12	42.8	79.3	2037	11	AY781407	AY781407 Synthetic
13	42.8	79.3	2062	11	AY781402	AY781402 Synthetic
c 14	42.8	79.3	2062	11	AY781402	AY781402 Synthetic
15	42.8	79.3	2162	11	AY781406	AY781406 Synthetic
c 16	42.8	79.3	2162	11	AY781406	AY781406 Synthetic
17	42.8	79.3	2413	11	AY781399	AY781399 Synthetic
c 18	42.8	79.3	2413	11	AY781399	AY781399 Synthetic
19	42.8	79.3	2933	11	AY781400	AY781400 Synthetic
c 20	42.8	79.3	2933	11	AY781400	AY781400 Synthetic
21	42.8	79.3	4791	11	AY738638	AY738638 Cloning v
c 22	42.8	79.3	4791	11	AY738638	AY738638 Cloning v
23	40.8	75.6	120	11	SYNGENE	M12560 Synthetic g
c 24	39.8	73.7	324	11	SYNPMUEND	M15949 Synthetic B
c 25	39.4	73.0	58	6	AR084420	AR084420 Sequence
c 26	39.4	73.0	71	6	AR084419	AR084419 Sequence
27	39.4	73.0	77	6	BD073246	BD073246 In vitro
28	39.4	73.0	77	6	AR353866	AR353866 Sequence
c 29	39.4	73.0	82	7	PMUBSR	M10177 Bacteriopha
c 30	39.4	73.0	83	6	AR084418	AR084418 Sequence
c 31	39.4	73.0	117	6	AR084417	AR084417 Sequence
32	39.4	73.0	150	7	PMUSEE1	M10863 Bacteriopha
c 33	39.4	73.0	220	7	NCMU3R	X05582 Mu-derived
c 34	39.4	73.0	220	7	PMUNE2	M34920 Bacteriopha
c 35	39.4	73.0	240	1	ECOTRPAA	M33723 Escherichia
36	39.4	73.0	903	6	A02708	A02708 pNM506 DNA
37	39.4	73.0	4666	1	ECOPOLBDA	M35371 E.coli DNA
c 38	39.4	73.0	7003	6	AR084426	AR084426 Sequence
39	39.4	73.0	15611	1	ECOPHNAQ	J05260 E.coli psiD
40	39.4	73.0	15611	6	AR229537	AR229537 Sequence
c 41	39.4	73.0	36717	7	AF083977	AF083977 Bacteriop
c 42	39.4	73.0	37199	11	AY860420	AY860420 Cloning v
43	37.4	69.3	240	1	ECOTRPAA	M33723 Escherichia
44	31.8	58.9	5755	1	ECOCYS	M32101 E.coli thio
c 45	30.4	56.3	1925	1	ECOMALIXA	M28539 E.coli malI

ALIGNMENTS

RESULT 1
AY781404
LOCUS AY781404 1319 bp DNA linear SYN 28-FEB-2005
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR6, complete sequence.
ACCESSION AY781404
VERSION AY781404.1 GI:60171997
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1319)
AUTHORS Zhang,C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street, Clayton, Victoria 3168, Australia
FEATURES Location/Qualifiers
source 1. .1319
/organism="synthetic construct"
/mol_type="other DNA"
/db_xref="taxon:32630"
repeat_region 1. .1319
/transposon="mini-Mu transposon TnCR6"
CDS 300. .959
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="AAX14471.1"
/db_xref="GI:60171998"
/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
A"
ORIGIN

Query Match 79.3%; Score 42.8; DB 11; Length 1319;
Best Local Similarity 87.0%; Pred. No. 0.00055;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54

```
              |||||||      | ||||||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 2
AY781404/c
LOCUS       AY781404               1319 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR6, complete
            sequence.
ACCESSION   AY781404
VERSION     AY781404.1  GI:60171997
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1319)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1319
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1319
                     /transposon="mini-Mu transposon TnCR6"
     CDS             300. .959
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAX14471.1"
                     /db_xref="GI:60171998"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1319;
  Best Local Similarity   87.0%;  Pred. No. 0.00055;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||      | ||||||||||||||||||||||||||||||||||||||||
Db       1318 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1265


RESULT 3
AY781405
LOCUS       AY781405               1351 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR7, complete
            sequence.
ACCESSION   AY781405
VERSION     AY781405.1  GI:60172026
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1351)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 1351)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1351
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1351
                     /transposon="mini-Mu transposon TnCR7"
     CDS             321. .980
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAX14472.1"
                     /db_xref="GI:60172027"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1351;
  Best Local Similarity   87.0%;  Pred. No. 0.00055;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||      | ||||||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 4
AY781405/c
LOCUS       AY781405               1351 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR7, complete
            sequence.
ACCESSION   AY781405
VERSION     AY781405.1  GI:60172026
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1351)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 1351)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1351
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1351
                     /transposon="mini-Mu transposon TnCR7"
     CDS             321. .980
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAX14472.1"
                     /db_xref="GI:60172027"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1351;
  Best Local Similarity   87.0%;  Pred. No. 0.00055;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;
```

```
Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1350 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1297


RESULT 5
AY781401
LOCUS       AY781401               1524 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR4, complete
            sequence.
ACCESSION   AY781401
VERSION     AY781401.1  GI:60171909
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1524)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 1524)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1524
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1524
                     /transposon="mini-Mu transposon TnCR4"
     CDS             168. .1358
                     /codon_start=1
                     /transl_table=11
                     /product="tetracyclin resistance protein"
                     /protein_id="AAX14468.1"
                     /db_xref="GI:60171910"
                     /translation="MKSNNALIVILGTVTLDAVGIGLVMPVLPGLLRDIVHSDSIASH
                     YGVLLALYALMQFLCAPVLGALSDRFGRRPVLLASLLGATIDYAIMATTPVLWILYAG
                     RIVAGITGATGAVAGAYIADITDGEDRARHFGLMSACFGVGMVAGPVAGGLLGAISLH
                     APFLAAAVLNGLNLLLGCFLMQESHKGERRPMPLRAFNPVSSFRWARGMTIVAALMTV
                     FFIMQLVGQVPAALWVIFGEDRFRWSATMIGLSLAVFGILHALAQAFVTGPATKRFGE
                     KQAIIAGMAADALGYVLLAFATRGWMAFPIMILLASGGIGMPALQAMLSRQVDDDHQG
                     QLQGSLAALTSLTSITGPLIVTAIYAASASTWNGLAWIVGAALYLVCLPALRRGAWSR
                     ATST"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1524;
  Best Local Similarity   87.0%;  Pred. No. 0.00054;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 6
AY781401/c
LOCUS       AY781401               1524 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR4, complete
            sequence.
ACCESSION   AY781401
VERSION     AY781401.1  GI:60171909
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1524)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 1524)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1524
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1524
                     /transposon="mini-Mu transposon TnCR4"
     CDS             168. .1358
                     /codon_start=1
                     /transl_table=11
                     /product="tetracyclin resistance protein"
                     /protein_id="AAX14468.1"
                     /db_xref="GI:60171910"
                     /translation="MKSNNALIVILGTVTLDAVGIGLVMPVLPGLLRDIVHSDSIASH
                     YGVLLALYALMQFLCAPVLGALSDRFGRRPVLLASLLGATIDYAIMATTPVLWILYAG
                     RIVAGITGATGAVAGAYIADITDGEDRARHFGLMSACFGVGMVAGPVAGGLLGAISLH
                     APFLAAAVLNGLNLLLGCFLMQESHKGERRPMPLRAFNPVSSFRWARGMTIVAALMTV
                     FFIMQLVGQVPAALWVIFGEDRFRWSATMIGLSLAVFGILHALAQAFVTGPATKRFGE
                     KQAIIAGMAADALGYVLLAFATRGWMAFPIMILLASGGIGMPALQAMLSRQVDDDHQG
                     QLQGSLAALTSLTSITGPLIVTAIYAASASTWNGLAWIVGAALYLVCLPALRRGAWSR
                     ATST"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1524;
  Best Local Similarity   87.0%;  Pred. No. 0.00054;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       1523 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1470


RESULT 7
AY781408
LOCUS       AY781408               1808 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR1, complete
            sequence.
ACCESSION   AY781408
VERSION     AY781408.1  GI:60172112
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1808)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 1808)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1808
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1808
                     /transposon="mini-Mu transposon TnCR1"
     CDS             825. .1484
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
```

```
                     /protein_id="AAX14475.1"
                     /db_xref="GI:60172113"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1808;
  Best Local Similarity   87.0%;  Pred. No. 0.00053;
  Matches    47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 8
AY781408/c
LOCUS       AY781408               1808 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR1, complete
            sequence.
ACCESSION   AY781408
VERSION     AY781408.1  GI:60172112
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1808)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 1808)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1808
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1808
                     /transposon="mini-Mu transposon TnCR1"
     CDS             825. .1484
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAX14475.1"
                     /db_xref="GI:60172113"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1808;
  Best Local Similarity   87.0%;  Pred. No. 0.00053;
  Matches    47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db       1807 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1754


RESULT 9
AY781403
LOCUS       AY781403               1937 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR5B, complete
            sequence.
ACCESSION   AY781403
VERSION     AY781403.1  GI:60171960
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1937)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1937
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1937
                     /transposon="mini-Mu transposon TnCR5B"
     CDS             607. .1401
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14470.1"
                     /db_xref="GI:60171961"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1937;
  Best Local Similarity   87.0%;  Pred. No. 0.00053;
  Matches    47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 10
AY781403/c
LOCUS       AY781403               1937 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR5B, complete
            sequence.
ACCESSION   AY781403
VERSION     AY781403.1  GI:60171960
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1937)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1937
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1937
                     /transposon="mini-Mu transposon TnCR5B"
     CDS             607. .1401
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14470.1"
                     /db_xref="GI:60171961"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
```

```
                EHQGLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 1937;
  Best Local Similarity   87.0%;  Pred. No. 0.00053;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db       1936 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1883


RESULT 11
AY781407
LOCUS       AY781407               2037 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR8B, complete
            sequence.
ACCESSION   AY781407
VERSION     AY781407.1  GI:60172085
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2037)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 2037)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .2037
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .2037
                     /transposon="mini-Mu transposon TnCR8B"
     CDS             644. .1438
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14474.1"
                     /db_xref="GI:60172086"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 2037;
  Best Local Similarity   87.0%;  Pred. No. 0.00052;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 12
AY781407/c
LOCUS       AY781407               2037 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR8B, complete
            sequence.
ACCESSION   AY781407
VERSION     AY781407.1  GI:60172085
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2037)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
   PUBMED   15699181
REFERENCE   2  (bases 1 to 2037)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .2037
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .2037
                     /transposon="mini-Mu transposon TnCR8B"
     CDS             644. .1438
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14474.1"
                     /db_xref="GI:60172086"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKASMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 2037;
  Best Local Similarity   87.0%;  Pred. No. 0.00052;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db       2036 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1983


RESULT 13
AY781402
LOCUS       AY781402               2062 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR5A, complete
            sequence.
ACCESSION   AY781402
VERSION     AY781402.1  GI:60171934
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2062)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .2062
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .2062
                     /transposon="mini-Mu transposon TnCR5A"
     CDS             732. .1526
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14469.1"
                     /db_xref="GI:60171935"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
```

```
                    LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                    EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                    QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 2062;
  Best Local Similarity   87.0%;  Pred. No. 0.00052;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 14
AY781402/c
LOCUS       AY781402               2062 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR5A, complete
            sequence.
ACCESSION   AY781402
VERSION     AY781402.1  GI:60171934
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2062)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .2062
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .2062
                     /transposon="mini-Mu transposon TnCR5A"
     CDS             732. .1526
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14469.1"
                     /db_xref="GI:60171935"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 2062;
  Best Local Similarity   87.0%;  Pred. No. 0.00052;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db       2061 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 2008


RESULT 15
AY781406
LOCUS       AY781406               2162 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR8A, complete
            sequence.
ACCESSION   AY781406
VERSION     AY781406.1  GI:60172057
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 2162)
  AUTHORS   Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
  TITLE     Transposon-mediated generation of targeting vectors for the
            production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
  PUBMED    15699181
REFERENCE   2  (bases 1 to 2162)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .2162
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .2162
                     /transposon="mini-Mu transposon TnCR8A"
     CDS             769. .1563
                     /codon_start=1
                     /transl_table=11
                     /product="neomycin phosphotransferase"
                     /protein_id="AAX14473.1"
                     /db_xref="GI:60172058"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
                     PVLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
                     LSSHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDE
                     EHQGLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
                     QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
ORIGIN

  Query Match             79.3%;  Score 42.8;  DB 11;  Length 2162;
  Best Local Similarity   87.0%;  Pred. No. 0.00052;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db          2 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


Search completed: January 18, 2006, 04:25:12
Job time : 280.233 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 33.5794 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-1
Perfect score: 54
Sequence: 1 gatctgattgattgaacgaa.....ttcacgataaatgcgaaaac 54

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	54	100.0	54	10	ACF58168	Acf58168 Cat-Mu tr
	2	54	100.0	1254	10	ACF58169	Acf58169 Modified
c	3	54	100.0	1254	10	ACF58169	Acf58169 Modified
	4	50	92.6	50	10	ACF58172	Acf58172 Cat-Mu tr
	5	42.8	79.3	54	4	AAD21279	Aad21279 Precut tr
	6	42.8	79.3	54	10	ACF58170	Acf58170 Cat-Mu tr
	7	42.8	79.3	82	10	ACC80745	Acc80745 Transposo
	8	42.8	79.3	83	10	ACC80741	Acc80741 Transposo
	9	42.8	79.3	84	10	ACC80738	Acc80738 Transposo
	10	42.8	79.3	86	10	ACC80742	Acc80742 Transposo
	11	42.8	79.3	89	10	ACC80753	Acc80753 Transposo
	12	41.2	76.3	54	10	ACF58171	Acf58171 Cat-Mu(No
c	13	39.4	73.0	51	13	ADS13821	Ads13821 Mu end DN
	14	39.4	73.0	56	13	ADS13820	Ads13820 Mu end DN
c	15	39.4	73.0	58	2	AAZ28883	Aaz28883 Mini-Mu t
c	16	39.4	73.0	71	2	AAZ28882	Aaz28882 Mini-Mu t
c	17	39.4	73.0	83	2	AAZ28881	Aaz28881 Mini-Mu t
	18	39.4	73.0	90	13	ADR14827	Adr14827 PCR prime
c	19	39.4	73.0	117	2	AAZ28880	Aaz28880 Mini-Mu t
	20	39.4	73.0	903	1	AAN80102	Aan80102 EcoRI/Bam
c	21	39.4	73.0	7003	2	AAZ28877	Aaz28877 Transposa
	22	39.4	73.0	15611	3	AAD01008	Aad01008 Escherich
	23	28	51.9	30	4	AAD21278	Aad21278 Bacteriop
c	24	25.8	47.8	1874	6	AAS98204	Aas98204 DNA encod
c	25	25.8	47.8	2229	7	ADS73196	Ads73196 Human kid
c	26	25.8	47.8	2229	7	ADW42050	Adw42050 cDNA elev
c	27	25.8	47.8	6136	6	ABL70342	Abl70342 Chemicall
c	28	25.8	47.8	6136	6	AAS61297	Aas61297 Human gen
c	29	25.8	47.8	6730	11	ADL22564	Adl22564 Human dis
c	30	25.8	47.8	17687	4	AAK71665	Aak71665 Human imm
c	31	25.8	47.8	17687	4	AAK64966	Aak64966 Human imm
c	32	25.8	47.8	17979	4	AAK71664	Aak71664 Human imm
c	33	25.8	47.8	17979	4	AAK64964	Aak64964 Human imm
c	34	25	46.3	34	14	AEB45577	Aeb45577 SigA2 wit
c	35	25	46.3	34	14	AEB48763	Aeb48763 Beta-lact
c	36	24.4	45.2	2000	11	ACL38625	Acl38625 Rice stre
	37	24.2	44.8	1995	10	ADE63262	Ade63262 Rat gene
	38	24.2	44.8	1995	10	ADD46332	Add46332 Rat gene
c	39	24	44.4	2099	6	ABQ69021	Abq69021 Listeria
c	40	24	44.4	7868	6	ABQ71053	Abq71053 Listeria
c	41	24	44.4	110000	6	ABQ69245_09	Continuation (10 o
c	42	24	44.4	110000	6	ABQ67197_08	Continuation (9 of
	43	23.8	44.1	2022	13	ADR85671	Adr85671 Aspergill
	44	23.8	44.1	2022	13	ADR85084	Adr85084 Aspergill
c	45	23.8	44.1	4861	4	ABL29670	Abl29670 Drosophil

ALIGNMENTS

RESULT 1
ACF58168
ID ACF58168 standard; DNA; 54 BP.
XX
AC ACF58168;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu transposon modified end fragment.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .4
FT /*tag= a
FT /note= "the 5' end of this strand overhangs the 3'end of
FT the complementary strand"
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Claim 6; Fig 2; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a

```
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a Cat-Mu transposon modified end
CC   fragment
XX
SQ   Sequence 54 BP; 22 A; 9 C; 12 G; 11 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 54;  DB 10;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-10;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 2
ACF58169
ID   ACF58169 standard; DNA; 1254 BP.
XX
AC   ACF58169;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Modified Cat-Mu(Stop)-transposon.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Synthetic.
OS   Bacteriophage mu.
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Claim 6; Page 30; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
SQ   Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 54;  DB 10;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-10;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 3
ACF58169/c
ID   ACF58169 standard; DNA; 1254 BP.
XX
AC   ACF58169;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Modified Cat-Mu(Stop)-transposon.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Synthetic.
OS   Bacteriophage mu.
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Claim 6; Page 30; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
SQ   Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 54;  DB 10;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-10;
  Matches   54;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1254 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1201


RESULT 4
ACF58172
ID   ACF58172 standard; DNA; 50 BP.
XX
```

```
AC   ACF58172;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Cat-Mu transposon modified end fragment.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Synthetic.
OS   Bacteriophage mu.
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Claim 6; Page 31; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a Cat-Mu transposon modified end
CC   fragment without 5' overhang
XX
SQ   Sequence 50 BP; 21 A; 8 C; 11 G; 10 T; 0 U; 0 Other;

  Query Match             92.6%;  Score 50;  DB 10;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-08;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          5 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50


RESULT 5
AAD21279
ID   AAD21279 standard; DNA; 54 BP.
XX
AC   AAD21279;
XX
DT   11-SEP-2003  (revised)
DT   28-JAN-2002  (first entry)
XX
DE   Precut transposon end of Bacteriophage Mu non-transferred strand.
XX
KW   Insertional mutation; synaptic complex; transposon; screening; ds.
XX
OS   Enterobacteria phage Mu.
XX
PN   US6294385-B1.
XX
PD   25-SEP-2001.
XX
PF   10-AUG-2000; 2000US-00635969.
XX
PR   23-SEP-1998;   98US-00159363.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Goryshin IY,  Reznikoff WS;
XX
DR   WPI; 2001-656171/75.
XX
PT   Making an insertional mutations, especially useful for efficiently
PT   inserting a transposable polynucleotide in a target cell, comprises
PT   introducing into the target cell a synaptic complex.
XX
PS   Claim 6; Col 2; 11pp; English.
XX
CC   The present invention relates to a method for making an insertional
CC   mutation at a random or quasi-random position in cellular nucleic acid in
CC   a target cell comprising introducing into the target cell a synaptic
CC   complex. The method is particularly useful for efficiently inserting a
CC   transposable polynucleotide at random or quasi-random locations in the
CC   chromosomal or extra-chromosomal nucleic acid of a target cell. The
CC   method may also be used for screening the genome of cells that comprise
CC   an insertional mutation that induces a phenotypic or genotypic change
CC   relative to the cells that are not subject to insertional mutagenesis.
CC   The present sequence is the precut transposon end of Bacteriophage Mu non
CC   -transferred strand, used in the exemplification of the invention.
CC   (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ   Sequence 54 BP; 21 A; 12 C; 14 G; 7 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 42.8;  DB 4;  Length 54;
  Best Local Similarity   87.0%;  Pred. No. 6.4e-06;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||       | ||||||||||||||||||||||||||||||||||||
Db          1 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 6
ACF58170
ID   ACF58170 standard; DNA; 54 BP.
XX
AC   ACF58170;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Cat-Mu transposon containing wild-type Mu ends.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Bacteriophage mu.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .4
FT                   /*tag=  a
FT                   /note= "the 5' end of this strand overhangs the 3'end of
FT                   the complementary strand"
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
```

XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Example 4; Fig 2; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a
CC transposon containing a genetically engineered translation stop signal
CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a Cat-Mu transposon containing wild
CC -type Mu ends
XX
SQ Sequence 54 BP; 21 A; 12 C; 14 G; 7 T; 0 U; 0 Other;

Query Match 79.3%; Score 42.8; DB 10; Length 54;
Best Local Similarity 87.0%; Pred. No. 6.4e-06;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54

RESULT 7
ACC80745
ID ACC80745 standard; DNA; 82 BP.
XX
AC ACC80745;
XX
DT 15-OCT-2003 (first entry)
XX
DE Transposon-based targeting construct related primer Mu2-Neo-2.
XX
KW Targeting construct; targeting vector; transposon; recombination;
KW deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS Synthetic.
XX
PN WO2003031629-A1.
XX
PD 17-APR-2003.
XX
PF 08-OCT-2002; 2002WO-AU001367.
XX
PR 09-OCT-2001; 2001AU-00008174.
PR 23-MAY-2002; 2002AU-00002522.
XX
PA (COPY-) COPYRAT PTY LTD.
XX
PI Morrison J, Zhang C;
XX
DR WPI; 2003-393445/37.
XX
PT Preparing a targeting construct using a transposon and DNA recombination
PT sequence, useful in making a targeting vector capable of modifying plant
PT and/or animal genome in a predetermined way.
XX
PS Example 5; Page 51; 92pp; English.
XX
CC The invention relates to methods of preparing a targeting construct for
CC use in a targeting vector capable of modifying a target DNA sequence, by
CC obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC sequence comprising a transposon sequence and a DNA recombination
CC sequence at two sites in the copy of the target DNA sequence, and
CC inducing a recombination event between the recombination sequences to
CC delete a portion of the copy of the target DNA sequence. The methods and
CC compositions of the present invention are useful for preparing a target
CC construct for use in a targeting vector for gene targeting or homologous
CC recombination. They can also be used for precisely modifying plant and/or
CC animal genome in a predetermined way. This sequence represents a primer
CC used in an example of the invention
XX
SQ Sequence 82 BP; 35 A; 21 C; 16 G; 10 T; 0 U; 0 Other;

Query Match 79.3%; Score 42.8; DB 10; Length 82;
Best Local Similarity 87.0%; Pred. No. 6.8e-06;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 61

RESULT 8
ACC80741
ID ACC80741 standard; DNA; 83 BP.
XX
AC ACC80741;
XX
DT 15-OCT-2003 (first entry)
XX
DE Transposon-based targeting construct related primer Mu1-4.
XX
KW Targeting construct; targeting vector; transposon; recombination;
KW deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS Synthetic.
XX
PN WO2003031629-A1.
XX
PD 17-APR-2003.
XX
PF 08-OCT-2002; 2002WO-AU001367.
XX
PR 09-OCT-2001; 2001AU-00008174.
PR 23-MAY-2002; 2002AU-00002522.
XX
PA (COPY-) COPYRAT PTY LTD.
XX
PI Morrison J, Zhang C;
XX
DR WPI; 2003-393445/37.
XX
PT Preparing a targeting construct using a transposon and DNA recombination
PT sequence, useful in making a targeting vector capable of modifying plant
PT and/or animal genome in a predetermined way.
XX
PS Example 5; Page 50; 92pp; English.
XX
CC The invention relates to methods of preparing a targeting construct for
CC use in a targeting vector capable of modifying a target DNA sequence, by
CC obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC sequence comprising a transposon sequence and a DNA recombination
CC sequence at two sites in the copy of the target DNA sequence, and
CC inducing a recombination event between the recombination sequences to
CC delete a portion of the copy of the target DNA sequence. The methods and
CC compositions of the present invention are useful for preparing a target
CC construct for use in a targeting vector for gene targeting or homologous
CC recombination. They can also be used for precisely modifying plant and/or
CC animal genome in a predetermined way. This sequence represents a primer
CC used in an example of the invention
XX
SQ Sequence 83 BP; 29 A; 20 C; 18 G; 16 T; 0 U; 0 Other;

Query Match 79.3%; Score 42.8; DB 10; Length 83;

```
  Best Local Similarity   87.0%;  Pred. No. 6.8e-06;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||      | ||||||||||||||||||||||||||||||||||||
Db         11 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 64


RESULT 9
ACC80738
ID   ACC80738 standard; DNA; 84 BP.
XX
AC   ACC80738;
XX
DT   15-OCT-2003  (first entry)
XX
DE   Transposon-based targeting construct related primer Mu1-1.
XX
KW   Targeting construct; targeting vector; transposon; recombination;
KW   deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS   Synthetic.
XX
PN   WO2003031629-A1.
XX
PD   17-APR-2003.
XX
PF   08-OCT-2002; 2002WO-AU001367.
XX
PR   09-OCT-2001; 2001AU-00008174.
PR   23-MAY-2002; 2002AU-00002522.
XX
PA   (COPY-) COPYRAT PTY LTD.
XX
PI   Morrison J,  Zhang C;
XX
DR   WPI; 2003-393445/37.
XX
PT   Preparing a targeting construct using a transposon and DNA recombination
PT   sequence, useful in making a targeting vector capable of modifying plant
PT   and/or animal genome in a predetermined way.
XX
PS   Example 5; Page 50; 92pp; English.
XX
CC   The invention relates to methods of preparing a targeting construct for
CC   use in a targeting vector capable of modifying a target DNA sequence, by
CC   obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC   sequence comprising a transposon sequence and a DNA recombination
CC   sequence at two sites in the copy of the target DNA sequence, and
CC   inducing a recombination event between the recombination sequences to
CC   delete a portion of the copy of the target DNA sequence. The methods and
CC   compositions of the present invention are useful for preparing a target
CC   construct for use in a targeting vector for gene targeting or homologous
CC   recombination. They can also be used for precisely modifying plant and/or
CC   animal genome in a predetermined way. This sequence represents a primer
CC   used in an example of the invention
XX
SQ   Sequence 84 BP; 29 A; 20 C; 19 G; 16 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 42.8;  DB 10;  Length 84;
  Best Local Similarity   87.0%;  Pred. No. 6.8e-06;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||      | ||||||||||||||||||||||||||||||||||||
Db         11 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 64


RESULT 10
ACC80742
ID   ACC80742 standard; DNA; 86 BP.
XX
AC   ACC80742;
XX
DT   15-OCT-2003  (first entry)
XX
DE   Transposon-based targeting construct related primer Mu2-1.
XX
KW   Targeting construct; targeting vector; transposon; recombination;
KW   deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS   Synthetic.
XX
PN   WO2003031629-A1.
XX
PD   17-APR-2003.
XX
PF   08-OCT-2002; 2002WO-AU001367.
XX
PR   09-OCT-2001; 2001AU-00008174.
PR   23-MAY-2002; 2002AU-00002522.
XX
PA   (COPY-) COPYRAT PTY LTD.
XX
PI   Morrison J,  Zhang C;
XX
DR   WPI; 2003-393445/37.
XX
PT   Preparing a targeting construct using a transposon and DNA recombination
PT   sequence, useful in making a targeting vector capable of modifying plant
PT   and/or animal genome in a predetermined way.
XX
PS   Example 5; Page 50; 92pp; English.
XX
CC   The invention relates to methods of preparing a targeting construct for
CC   use in a targeting vector capable of modifying a target DNA sequence, by
CC   obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC   sequence comprising a transposon sequence and a DNA recombination
CC   sequence at two sites in the copy of the target DNA sequence, and
CC   inducing a recombination event between the recombination sequences to
CC   delete a portion of the copy of the target DNA sequence. The methods and
CC   compositions of the present invention are useful for preparing a target
CC   construct for use in a targeting vector for gene targeting or homologous
CC   recombination. They can also be used for precisely modifying plant and/or
CC   animal genome in a predetermined way. This sequence represents a primer
CC   used in an example of the invention
XX
SQ   Sequence 86 BP; 27 A; 20 C; 20 G; 19 T; 0 U; 0 Other;

  Query Match             79.3%;  Score 42.8;  DB 10;  Length 86;
  Best Local Similarity   87.0%;  Pred. No. 6.8e-06;
  Matches   47;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||||||      | ||||||||||||||||||||||||||||||||||||
Db         11 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 64


RESULT 11
ACC80753
ID   ACC80753 standard; DNA; 89 BP.
XX
AC   ACC80753;
XX
DT   15-OCT-2003  (first entry)
XX
DE   Transposon-based targeting construct related primer Mu-polyA-2.
XX
KW   Targeting construct; targeting vector; transposon; recombination;
KW   deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS   Synthetic.
XX
PN   WO2003031629-A1.
XX
PD   17-APR-2003.
```

XX
PF 08-OCT-2002; 2002WO-AU001367.
XX
PR 09-OCT-2001; 2001AU-00008174.
PR 23-MAY-2002; 2002AU-00002522.
XX
PA (COPY-) COPYRAT PTY LTD.
XX
PI Morrison J, Zhang C;
XX
DR WPI; 2003-393445/37.
XX
PT Preparing a targeting construct using a transposon and DNA recombination
PT sequence, useful in making a targeting vector capable of modifying plant
PT and/or animal genome in a predetermined way.
XX
PS Example 5; Page 52; 92pp; English.
XX
CC The invention relates to methods of preparing a targeting construct for
CC use in a targeting vector capable of modifying a target DNA sequence, by
CC obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC sequence comprising a transposon sequence and a DNA recombination
CC sequence at two sites in the copy of the target DNA sequence, and
CC inducing a recombination event between the recombination sequences to
CC delete a portion of the copy of the target DNA sequence. The methods and
CC compositions of the present invention are useful for preparing a target
CC construct for use in a targeting vector for gene targeting or homologous
CC recombination. They can also be used for precisely modifying plant and/or
CC animal genome in a predetermined way. This sequence represents a primer
CC used in an example of the invention
XX
SQ Sequence 89 BP; 28 A; 19 C; 20 G; 22 T; 0 U; 0 Other;

Query Match 79.3%; Score 42.8; DB 10; Length 89;
Best Local Similarity 87.0%; Pred. No. 6.8e-06;
Matches 47; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
||||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 8 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 61

RESULT 12
ACF58171
ID ACF58171 standard; DNA; 54 BP.
XX
AC ACF58171;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu(NotI) transposon modified end fragment.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .4
FT /*tag= a
FT /note= "the 5' end of this strand overhangs the 3'end of
FT the complementary strand"
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Example 4; Fig 2; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a
CC transposon containing a genetically engineered translation stop signal
CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a modified Cat-Mu transposon
CC containing Mu ends with engineered NotI restriction site
XX
SQ Sequence 54 BP; 19 A; 14 C; 14 G; 7 T; 0 U; 0 Other;

Query Match 76.3%; Score 41.2; DB 10; Length 54;
Best Local Similarity 85.2%; Pred. No. 2.4e-05;
Matches 46; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
|||||| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 GATCTGCGGCCGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54

RESULT 13
ADS13821/c
ID ADS13821 standard; DNA; 51 BP.
XX
AC ADS13821;
XX
DT 16-DEC-2004 (first entry)
XX
DE Mu end DNA fragment construction oligonucleotide MM1141.
XX
KW ss; mismatch detection; transposition detection; pathogen typing;
KW embryo screening; mutation detection; Mu end DNA; MM1141.
XX
OS Enterobacteria phage Mu.
OS Synthetic.
XX
PN US2004191821-A1.
XX
PD 30-SEP-2004.
XX
PF 26-MAR-2004; 2004US-00809688.
XX
PR 28-MAR-2003; 2003US-0457934P.
XX
PA (USSA) US SEC OF ARMY.
XX
PI Yanagihara K, Mizuuchi K;
XX
DR WPI; 2004-689846/67.
XX
PT Detecting a mismatch in a test double stranded nucleic acid target,
PT useful for typing a pathogenic microorganism strain, comprises detecting
PT transposition of the Mu-end nucleic acid into the target.
XX
PS Example 1; SEQ ID NO 2; 24pp; English.
XX
CC The invention relates to a method of detecting a mismatch in a test
CC double stranded nucleic acid target which comprises detecting

CC transposition of the Mu-end nucleic acid into the target, where
CC transposition at the predominant site indicates the presence of mismatch
CC at about that site. The methods are useful for typing a pathogenic
CC microorganism strain, for screening an embryo for the presence of
CC mutation, for detecting the presence of known mutation in a gene of
CC interest, detecting the presence of a previously unidentified mutation in
CC a gene of interest, and diagnosing the presence or absence of a tumour-
CC promoting mutation. The kit is useful for detecting the presence of a
CC mutation or polymorphism of interest in a nucleic acid molecule. The
CC present sequence represents the Mu end DNA fragment construction
CC oligonucleotide MM1141.
XX
SQ Sequence 51 BP; 6 A; 13 C; 11 G; 21 T; 0 U; 0 Other;

Query Match 73.0%; Score 39.4; DB 13; Length 51;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
|||||||||||||||||||||||||||||||||||||||
Db 42 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 2

RESULT 14
ADS13820
ID ADS13820 standard; DNA; 56 BP.
XX
AC ADS13820;
XX
DT 16-DEC-2004 (first entry)
XX
DE Mu end DNA fragment construction oligonucleotide MM1138.
XX
KW ss; mismatch detection; transposition detection; pathogen typing;
KW embryo screening; mutation detection; Mu end DNA; MM1138.
XX
OS Enterobacteria phage Mu.
OS Synthetic.
XX
PN US2004191821-A1.
XX
PD 30-SEP-2004.
XX
PF 26-MAR-2004; 2004US-00809688.
XX
PR 28-MAR-2003; 2003US-0457934P.
XX
PA (USSA) US SEC OF ARMY.
XX
PI Yanagihara K, Mizuuchi K;
XX
DR WPI; 2004-689846/67.
XX
PT Detecting a mismatch in a test double stranded nucleic acid target,
PT useful for typing a pathogenic microorganism strain, comprises detecting
PT transposition of the Mu-end nucleic acid into the target.
XX
PS Example 1; SEQ ID NO 1; 24pp; English.
XX
CC The invention relates to a method of detecting a mismatch in a test
CC double stranded nucleic acid target which comprises detecting
CC transposition of the Mu-end nucleic acid into the target, where
CC transposition at the predominant site indicates the presence of mismatch
CC at about that site. The methods are useful for typing a pathogenic
CC microorganism strain, for screening an embryo for the presence of
CC mutation, for detecting the presence of known mutation in a gene of
CC interest, detecting the presence of a previously unidentified mutation in
CC a gene of interest, and diagnosing the presence or absence of a tumour-
CC promoting mutation. The kit is useful for detecting the presence of a
CC mutation or polymorphism of interest in a nucleic acid molecule. The
CC present sequence represents the Mu end DNA fragment construction
CC oligonucleotide MM1138.
XX
SQ Sequence 56 BP; 22 A; 12 C; 15 G; 7 T; 0 U; 0 Other;

Query Match 73.0%; Score 39.4; DB 13; Length 56;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
|||||||||||||||||||||||||||||||||||||||
Db 15 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55

RESULT 15
AAZ28883/c
ID AAZ28883 standard; cDNA; 58 BP.
XX
AC AAZ28883;
XX
DT 15-FEB-2000 (first entry)
XX
DE Mini-Mu transposable element deletion region delta-1.
XX
KW Transposon Mu; transposable element; fusion protein; attachment site;
KW attL; attR; protein domain library; enzyme; accelerated evolution; ss.
XX
OS Synthetic.
XX
PN US5981177-A.
XX
PD 09-NOV-1999.
XX
PF 25-JAN-1995; 95US-00378548.
XX
PR 25-JAN-1995; 95US-00378548.
XX
PA (DEMI/) DEMIRJIAN D C.
PA (CASA/) CASADABAN M J.
PA (WEBE/) WEBER J M.
PA (GAIN/) GAINES G L.
XX
PI Casadaban MJ, Demirjian DC, Weber JM, Gaines GL;
XX
DR WPI; 1999-633307/54.
XX
PT Generating fusion proteins using transposable elements, useful for
PT development of a protein domain library and in the construction of multi-
PT functional enzymes.
XX
PS Example 2; Fig 2; 41pp; English.
XX
CC The invention relates to a Mu-like transposable element (I) used for
CC generating functional fusion proteins after insertion into a target DNA.
CC The Mu-like element comprises: (a) a left transposable element attachment
CC site attL and a right transposable element attachment site attR, where
CC attR is no more than 62 nucleotides long; (b) a site for insertion of an
CC exogenous DNA sequence encoding for a protein domain located between attL
CC and attR; (c) after insertion of the transposable element into a target
CC DNA sequence, a fusion mRNA sequence is transcribed originating either
CC from the target DNA on either side of the transposable element or from
CC inside the transposable element and continuing through the attachment
CC site sequences and into the protein coding region, resulting in a single
CC fusion open reading frame (ORF). The constructs are useful in the
CC development of a protein domain library, in the construction of multi-
CC functional enzymes and in the accelerated evolution of new enzymatic
CC activities. The sequences AAZ28880-Z28884 represent deletion mini-Mu
CC elements of the invention (encoded ORF - AAY55901-Y55906)
XX
SQ Sequence 58 BP; 8 A; 15 C; 13 G; 22 T; 0 U; 0 Other;

Query Match 73.0%; Score 39.4; DB 2; Length 58;
Best Local Similarity 97.6%; Pred. No. 0.00011;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54

```
                  | ||||||||||||||||||||||||||||||||||||||
Db         49 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 9
```

Search completed: January 17, 2006, 19:35:05
Job time : 35.5794 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 228.555 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-1
Perfect score: 54
Sequence: 1 gatctgattgattgaacgaa.....ttcacgataaatgcgaaaac 54

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	40.4	74.8	370	10	CZ569838	CZ569838 OB__Ba000
	2	39.4	73.0	226	9	CC742071	CC742071 ZMMBBb011
	3	39.4	73.0	589	10	CL858345	CL858345 OR_CBa008
	4	36.2	67.0	543	10	CZ821413	CZ821413 OC__Ba019
	5	36.2	67.0	688	10	CZ820900	CZ820900 OC__Ba019
	6	30.6	56.7	247	10	CZ821171	CZ821171 OC__Ba019
	7	27.4	50.7	582	3	BJ016030	BJ016030 BJ016030
c	8	26.6	49.3	757	3	BI953957	BI953957 HVSMEm001
c	9	26.6	49.3	802	10	DU032104	DU032104 13515 Tom
c	10	26.4	48.9	274	8	T31184	T31184 EST28439 Hu
c	11	26.4	48.9	771	3	BI953902	BI953902 HVSMEm001
c	12	26.2	48.5	215	3	BI953982	BI953982 HVSMEm001
c	13	26.2	48.5	771	3	BI953934	BI953934 HVSMEm001
c	14	26	48.1	384	6	CD113955	CD113955 ME1-0032P
	15	26	48.1	510	9	BH821089	BH821089 BACPP15-D
c	16	26	48.1	737	3	BI954000	BI954000 HVSMEm001
c	17	25.8	47.8	248	8	F05165	F05165 HSC02H021 n
c	18	25.8	47.8	304	1	AA481614	AA481614 aa35g09.r
c	19	25.8	47.8	403	1	AA262843	AA262843 zs24d05.r
c	20	25.8	47.8	535	5	C14413	C14413 C14413 Clon
c	21	25.8	47.8	561	1	AW954984	AW954984 EST367054
c	22	25.8	47.8	634	1	AW955329	AW955329 EST367399
	23	25.8	47.8	699	9	AZ571139	AZ571139 283PvD05
c	24	25.8	47.8	731	1	AL041260	AL041260 DKFZp434L
c	25	25.8	47.8	856	3	BI772228	BI772228 603056184
c	26	25.8	47.8	994	3	BM804853	BM804853 AGENCOURT
c	27	25.6	47.4	196	3	BI953931	BI953931 HVSMEm001
	28	25.6	47.4	705	3	BQ149276	BQ149276 NF088G08F
c	29	25.6	47.4	752	3	BI953878	BI953878 HVSMEm001
	30	25.6	47.4	973	10	DU005263	DU005263 300893 To
c	31	25.4	47.0	484	9	AZ152875	AZ152875 SP_0046_B
	32	25.4	47.0	614	7	CJ393910	CJ393910 CJ393910
	33	25.4	47.0	638	7	CJ417250	CJ417250 CJ417250
c	34	25.4	47.0	749	3	BI953959	BI953959 HVSMEm001
c	35	25.4	47.0	887	7	CN586362	CN586362 USDA-FP_1
	36	25.2	46.7	297	6	CF504687	CF504687 USDA-FP_1
	37	25.2	46.7	491	8	DN797614	DN797614 USDA-FP7A
c	38	25.2	46.7	518	9	BH684386	BH684386 BOHXP87TF
c	39	25.2	46.7	671	9	BZ833089	BZ833089 CH240_222
	40	25.2	46.7	700	8	DR400612	DR400612 TKN060E06
c	41	25.2	46.7	703	5	BW085322	BW085322 BW085322
	42	25.2	46.7	706	3	BM617600	BM617600 170006871
c	43	25.2	46.7	741	3	BI953998	BI953998 HVSMEm001
c	44	25.2	46.7	748	3	BI954023	BI954023 HVSMEm001
	45	25.2	46.7	776	6	CB292852	CB292852 UCRCS01_0

ALIGNMENTS

RESULT 1
CZ569838
LOCUS CZ569838 370 bp DNA linear GSS 20-JUN-2005
DEFINITION OB__Ba0003F02.r OB__Ba Oryza brachyantha genomic clone
OB__Ba0003F02 3', genomic survey sequence.
ACCESSION CZ569838
VERSION CZ569838.1 GI:68013639
KEYWORDS GSS.
SOURCE Oryza brachyantha
ORGANISM Oryza brachyantha
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 370)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Insert Length: 145 Std Error: 0.00
Plate: 0003 row: F column: 02
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .370
/organism="Oryza brachyantha"
/mol_type="genomic DNA"
/db_xref="taxon:4533"
/clone="OB__Ba0003F02"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OB__Ba"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

```
  Query Match            74.8%;  Score 40.4;  DB 10;  Length 370;
  Best Local Similarity  97.6%;  Pred. No. 0.00042;
  Matches   41;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         13 TGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              || |||||||||||||||||||||||||||||||||||||||
Db         85 TGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 126


RESULT 2
CC742071
LOCUS       CC742071                 226 bp    DNA     linear   GSS 25-JUN-2003
DEFINITION  ZMMBBb0115J07.f ZMMBBb Zea mays genomic clone ZMMBBb0115J07 5',
            genomic survey sequence.
ACCESSION   CC742071
VERSION     CC742071.1  GI:32194524
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 226)
  AUTHORS   Yu,Y., Kim,H.R., Hatfield,J., Soderlund,C., Bharti,A.K., Messing,J.
            and Wing,R.
  TITLE     Sequencing of the maize genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: M13r
            Plate: 0115  row: J  column: 07
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .226
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBb0115J07"
                     /lab_host="DH10B"
                     /clone_lib="ZMMBBb"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Zea mays L. ssp. mays"
ORIGIN

  Query Match            73.0%;  Score 39.4;  DB 9;  Length 226;
  Best Local Similarity  97.6%;  Pred. No. 0.00095;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              | |||||||||||||||||||||||||||||||||||||||
Db        182 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 222


RESULT 3
CL858345
LOCUS       CL858345                 589 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089D12.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089D12
            5', genomic survey sequence.
ACCESSION   CL858345
VERSION     CL858345.1  GI:51269584
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: D  column: 12
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .589
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089D12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match            73.0%;  Score 39.4;  DB 10;  Length 589;
  Best Local Similarity  97.6%;  Pred. No. 0.00098;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              | |||||||||||||||||||||||||||||||||||||||
Db        495 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 535


RESULT 4
CZ821413
LOCUS       CZ821413                 543 bp    DNA     linear   GSS 26-JUL-2005
DEFINITION  OC__Ba0199P13.r OC__Ba Oryza coarctata genomic clone OC__Ba0199P13
            3', genomic survey sequence.
ACCESSION   CZ821413
VERSION     CZ821413.1  GI:71261266
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
```

```
           BACKWARD: CAC TCA TTA GGC ACC CCA
           Plate: 0199  row: P  column: 13
           Seq primer: CAC TCA TTA GGC ACC CCA
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0199P13"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match            67.0%;  Score 36.2;  DB 10;  Length 543;
  Best Local Similarity  92.7%;  Pred. No. 0.014;
  Matches   38;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              | |||||||| |||||||||||||||||||| ||||||||||
Db        473 GCACGAAAAATGCGAAAGCGTTTCACGATACATGCGAAAAC 513


RESULT 5
CZ820900
LOCUS       CZ820900                688 bp    DNA     linear   GSS 26-JUL-2005
DEFINITION  OC__Ba0199D21.r OC__Ba Oryza coarctata genomic clone OC__Ba0199D21
            3', genomic survey sequence.
ACCESSION   CZ820900
VERSION     CZ820900.1  GI:71260753
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0199  row: D  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0199D21"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match            67.0%;  Score 36.2;  DB 10;  Length 688;
  Best Local Similarity  92.7%;  Pred. No. 0.014;
  Matches   38;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              | |||||||||||||| ||||| ||||||||||||||||||
Db        440 GCACGAAAAACGCGACAGCGTCTCACGATAAATGCGAAAAC 480


RESULT 6
CZ821171
LOCUS       CZ821171                247 bp    DNA     linear   GSS 26-JUL-2005
DEFINITION  OC__Ba0199J21.r OC__Ba Oryza coarctata genomic clone OC__Ba0199J21
            3', genomic survey sequence.
ACCESSION   CZ821171
VERSION     CZ821171.1  GI:71261024
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0199  row: J  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0199J21"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match            56.7%;  Score 30.6;  DB 10;  Length 247;
  Best Local Similarity  73.6%;  Pred. No. 1.5;
  Matches   39;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 ATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              || |||     || |  |||||  |||||||||||||| ||| |||| ||||||
Db        132 ATATGAAGCGGTGCATCAAAAATCCGAAAGCGTTTCATGATCAATGTGAAAAC 184


RESULT 7
BJ016030
LOCUS       BJ016030                582 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ016030 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA008A02 3',
            mRNA sequence.
ACCESSION   BJ016030
VERSION     BJ016030.1  GI:17375178
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
```

```
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA008A02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             50.7%;  Score 27.4;  DB 3;  Length 582;
  Best Local Similarity   69.8%;  Pred. No. 23;
  Matches   37;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAA 53
              | | | ||| ||| ||||  |||  | ||||| |||||  |||||| | || |
Db        206 GTTGTCATTTATTAAACGTGAAAACCAAAAGCATTTCAGAATAAATACTAACA 258


RESULT 8
BI953957/c
LOCUS       BI953957                757 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0015I12f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
            clone HVSMEm0015I12f, mRNA sequence.
ACCESSION   BI953957
VERSION     BI953957.1  GI:16299057
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
            Simmons,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 590
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 4
            High quality sequence stop: 756.
FEATURES             Location/Qualifiers
     source          1. .757
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEm0015I12f"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN

  Query Match             49.3%;  Score 26.6;  DB 3;  Length 757;
  Best Local Similarity   81.1%;  Pred. No. 46;
  Matches   43;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy          2 ATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              | ||||      | |||||||||||||||| ||||||| ||||||||||||
Db         52 AGCTGAAGCGGCGCACGAAAAACGCGAAAGCTTTTCACG-TAAATGCGAAAAC 1


RESULT 9
DU032104/c
LOCUS       DU032104                802 bp    DNA     linear   GSS 12-AUG-2005
DEFINITION  13515 Tomato HindIII BAC Library Lycopersicon esculentum genomic
            clone LE_HBa0169A20 3, genomic survey sequence.
ACCESSION   DU032104
VERSION     DU032104.1  GI:72445405
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 802)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 13518
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 169  row: A  column: 20
            Seq primer: SP6
```

```
            Class: BAC ends
            High quality sequence start: 36
            High quality sequence stop: 428.
FEATURES              Location/Qualifiers
     source           1. .802
                      /organism="Lycopersicon esculentum"
                      /mol_type="genomic DNA"
                      /cultivar="Heinz 1706"
                      /db_xref="taxon:4081"
                      /clone="LE_HBa0169A20"
                      /lab_host="E. coli"
                      /clone_lib="Tomato HindIII BAC Library"
                      /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             49.3%;  Score 26.6;  DB 10;  Length 802;
  Best Local Similarity   71.4%;  Pred. No. 46;
  Matches   35;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          6 GATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              |||  ||| || ||||||   | |||  |||||| | | ||| ||||||
Db        693 GATGAATTAAAAGAAAAAGAAGCAAGATTTTCACTAAATATGAGAAAAC 645


RESULT 10
T31184/c
LOCUS       T31184                  274 bp    mRNA    linear   EST 06-SEP-1995
DEFINITION  EST28439 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
            sequence.
ACCESSION   T31184
VERSION     T31184.1  GI:613282
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 274)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
            Bult,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
            White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W.,
            Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
            FitzGerald,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M.,
            Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S.,
            Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M.,
            Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T.,
            Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
            Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
            Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A.,
            Coleman,T.A., Collins,E.-J., Dimke,D., Feng,P., Ferrie,A.,
            Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,J.M.,
            Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Ji,H., Li,H.,
            Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wing,J., Xu,C.,
            Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
            Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
   PUBMED   7566098
COMMENT     Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .274
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="ATCC (inhost):101229"
                      /db_xref="taxon:9606"
                      /clone_lib="Human Brain"
                      /note="Organ: brain"
ORIGIN

  Query Match             48.9%;  Score 26.4;  DB 8;  Length 274;
  Best Local Similarity   67.9%;  Pred. No. 53;
  Matches   36;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          2 ATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54
              || |  || | | ||| |||||  | ||| | | |||| | ||||| || |||
Db         93 ATTTATTTCAATTAACTAAAAATTCCAAAACTTATCACNAAAAATGAGACAAC 41


RESULT 11
BI953902/c
LOCUS       BI953902                771 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0015E24f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
            clone HVSMEm0015E24f, mRNA sequence.
ACCESSION   BI953902
VERSION     BI953902.1  GI:16298957
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
            Simmons,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 486
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 16
            High quality sequence stop: 619.
FEATURES              Location/Qualifiers
     source           1. .771
                      /organism="Hordeum vulgare subsp. vulgare"
                      /mol_type="mRNA"
                      /cultivar="Morex"
                      /sub_species="vulgare"
                      /db_xref="taxon:112509"
                      /clone="HVSMEm0015E24f"
                      /tissue_type="green seedling leaf"
                      /lab_host="TJC121"
                      /clone_lib="Hordeum vulgare green seedling EST library
                      HVcDNA0014 (Blumeria infected)"
                      /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                      XhoI; Morex (mla) plants were greenhouse grown in the R
                      Wise lab at Iowa State University, Ames, IA; 7 day old
                      green seedlings were infected with isolate 5874 of
                      Blumeria graminis f. sp. hordei, and leaves were harvested
                      24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                      In the TJ Close lab at the University of California,
                      Riverside, total RNA was prepared from each sample pool,
                      equal quantities of all three RNA pools were combined,
                      poly(A) RNA was purified from the mixture, one primary
                      unamplified cDNA library was made, and 1 million pfu were
                      in vivo excised to give pBluescript SK(-) cDNA phagemids
                      (Chin). Phagemids were plated and picked at the Clemson
```

University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 48.9%; Score 26.4; DB 3; Length 771;
Best Local Similarity 75.0%; Pred. No. 54;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy          2 ATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAA 45
              || |||      | |||||||||||||||||| |||||||  |||
Db         44 ATGTGAAGCGGCGCACGAAAAACGCGAAAGCTTTTCACGTAAAA 1
```

RESULT 12
BI953982/c
LOCUS BI953982 215 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEm0015K12f Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA clone HVSMEm0015K12f, mRNA sequence.
ACCESSION BI953982
VERSION BI953982.1 GI:16299103
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 215)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 174
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 4
High quality sequence stop: 216.

FEATURES Location/Qualifiers
source 1. .215
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEm0015K12f"
/tissue_type="green seedling leaf"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Morex (mla) plants were greenhouse grown in the R Wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 48.5%; Score 26.2; DB 3; Length 215;
Best Local Similarity 79.5%; Pred. No. 62;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy         14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAA 52
              | ||||||||||||||||||| ||||  |||    |||
Db         40 GCACGAAAAACGCGAAAGCGTTGCACGTAAAAGCGAAAA 2
```

RESULT 13
BI953934/c
LOCUS BI953934 771 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEm0015G22f Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA clone HVSMEm0015G22f, mRNA sequence.
ACCESSION BI953934
VERSION BI953934.1 GI:16299017
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 771)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 471
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 25
High quality sequence stop: 736.

FEATURES Location/Qualifiers
source 1. .771
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"

/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEm0015G22f"
/tissue_type="green seedling leaf"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Morex (mla) plants were greenhouse grown in the R Wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 48.5%; Score 26.2; DB 3; Length 771;
Best Local Similarity 79.5%; Pred. No. 64;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy         14 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAA 52
              | ||||||||||||||||| |||||||  |||    |||
Db         40 GCACGAAAAACGCGAAAGCTTTTCACGTAAAAGCAAAAA 2
```

RESULT 14
CD113955/c
LOCUS CD113955 384 bp mRNA linear EST 14-SEP-2003
DEFINITION ME1-0032P-A101-F03-U.B ME1-0032 Schistosoma mansoni cDNA clone ME1-0032P-A101-F03.B, mRNA sequence.
ACCESSION CD113955
VERSION CD113955.1 GI:34652145
KEYWORDS EST.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 384)
AUTHORS Verjovski-Almeida,S., DeMarco,R., Martins,E.A.L., Guimaraes,P.E.M., Ojopi,E.P.B., Paquola,A.C.M., Piazza,J.P., Nishiyama,M.Y. Jr., Kitajima,J.P., Adamson,R.E., Ashton,P.D., Bonaldo,M.F., Coulson,P.S., Dillon,G.P., Farias,L.P., Gregorio,S.P., Ho,P.L., Leite,R.A., Malaquias,L.C.C., Marques,R.C.P., Miyasato,P.A., Nascimento,A.L.T.O., Ohlweiler,F.P., Reis,E.M., Ribeiro,M.A., Sa,R.G., Stukart,G.C., Soares,M.B., Gargioni,C., Kawano,T., Rodrigues,V., Madeira,A.M.B.N., Wilson,R.A., Menck,C.F.M., Setubal,J.C., Leite,L.C.C. and Dias-Neto,E.
TITLE Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni
JOURNAL Nat. Genet. 35 (2), 148-157 (2003)
PUBMED 12973350
COMMENT Contact: Dr. Sergio Verjovski-Almeida
Departamento de Bioquimica
Instituto de Quimica - Universidade de Sao Paulo
Av. Prof. Lineu Prestes 748 sala 1200, 05508-900 Sao Paulo - SP, Brasil
Tel: +55-11-3091-2173
Fax: +55-11-3091-2186
Email: verjo@iq.usp.br
This sequence was derived from the FAPESP Schistosoma mansoni EST Genome Project. All sequences in the project were assembled and annotated. This entry and all the assembled sequences can be seen in the following URL http://bioinfo.iq.usp.br/schisto/
Plate: ME1-0032P-A101 row: 3 column: F.
FEATURES Location/Qualifiers
source 1. .384
/organism="Schistosoma mansoni"
/mol_type="mRNA"
/db_xref="taxon:6183"
/clone="ME1-0032P-A101-F03.B"
/sex="mixed pool"
/dev_stage="egg"
/lab_host="Mus musculus"
/clone_lib="ME1-0032"
/note="Vector: pGEM T-easy"
ORIGIN

Query Match 48.1%; Score 26; DB 6; Length 384;
Best Local Similarity 70.0%; Pred. No. 74;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          2 ATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAA 51
              |||| ||  | ||||||||||| |  ||   |||| ||||   |||| ||
Db         50 ATCTCATCCAATGAACGAAAAATGAAAATTGGTTTAACGACTGATGCTAA 1
```

RESULT 15
BH821089
LOCUS BH821089 510 bp DNA linear GSS 20-MAY-2002
DEFINITION BACPP15-D03.y Pristionchus pacificus BAC ends Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION BH821089
VERSION BH821089.1 GI:20996539
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 510)
AUTHORS Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz,G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A., Jesse,T., Millare,J., de Both,M., Schuster,S.C. and Sommer,R.J.
TITLE A BAC-based genetic linkage map of the nematode Pristionchus pacificus
JOURNAL Unpublished (2002)
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .510
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="var. California"
/db_xref="taxon:54126"
/clone_lib="Pristionchus pacificus BAC ends"
ORIGIN

Query Match 48.1%; Score 26; DB 9; Length 510;
Best Local Similarity 70.0%; Pred. No. 75;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGA 50
              ||||  ||||||| || ||||| || | |||||  |    | ||||| ||
Db        253 GATCATATTGATTAAATGAAAATCGGGGAAGCGGATATTTAGAAATGAGA 302


Search completed: January 18, 2006, 11:40:45
Job time : 232.555 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 6461.19 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-2
Perfect score: 1254
Sequence: 1 gatctgattgattgaacgaa.....ttcgttcaatcaatcagatc 1254

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	1147.2	91.5	3400	6	AR477289	Sequence
c	2	1142.8	91.1	6122	6	AX554413	Sequence
c	3	1142.8	91.1	6122	6	AX554422	Sequence
c	4	1142.8	91.1	6122	6	AX642149	Sequence
c	5	1142.8	91.1	6125	6	AX554420	Sequence
c	6	1135.2	90.5	3399	11	CVU46018	U46018 Cloning vec
c	7	1135.2	90.5	4100	6	CS119871	Sequence
c	8	1131.8	90.3	1808	11	AY781408	AY781408 Synthetic
	9	1130.8	90.2	3417	11	AF153422	AF153422 Cloning v
c	10	1129.6	90.1	1851	1	ACCTN2670L	M37690 Acinetobact
c	11	1129.6	90.1	1874	1	ACCACEAA	M62822 A.baumannii
	12	1129.6	90.1	6464	1	AY608912	AY608912 Escherich
	13	1129.6	90.1	11925	11	AY303237	AY303237 Shuttle v
	14	1129.6	90.1	13515	11	AY303236	AY303236 Shuttle v
c	15	1129.6	90.1	45325	1	AY123253	AY123253 Klebsiell
	16	1129.6	90.1	94281	1	AP000342	AP000342 Plasmid R
	17	1129.6	90.1	120592	3	AJ851089	AJ851089 Unculture
c	18	1129.6	90.1	218160	1	STYPPHCM1	AL513383 Salmonell
c	19	1127.4	89.9	5593	11	AF519766	AF519766 Cloning v
c	20	1127.2	89.9	7830	11	AY428809	AY428809 Cloning v
	21	1125.2	89.7	2933	11	AY781400	AY781400 Synthetic
	22	1122.6	89.5	4376	1	SFU81140	U81140 Shigella fl
	23	1122.6	89.5	66715	1	AF326777	AF326777 Shigella
c	24	1121.8	89.5	1319	11	AY781404	AY781404 Synthetic
c	25	1118.8	89.2	8530	11	AY289598	AY289598 Cloning v
	26	1118.6	89.2	2701	11	AF060240	AF060240 Mobile an
c	27	1116.8	89.1	2839	11	AY222812	AY222812 Cloning v
c	28	1116.8	89.1	4152	11	AY219693	AY219693 Shuttle v
c	29	1116.8	89.1	5498	11	AY222821	AY222821 Shuttle v
	30	1116	89.0	3930	11	AB001326	AB001326 Cloning v
c	31	1116	89.0	5240	11	AB052891	AB052891 Cloning v
	32	1116	89.0	5273	11	EVU51556	U51556 Expression
	33	1116	89.0	6241	11	AF092036	AF092036 Shuttle v
	34	1116	89.0	6758	11	AF121784	AF121784 Expressio
	35	1116	89.0	8774	11	IVU69267	U69267 Integration
	36	1116	89.0	9065	11	AF405698	AF405698 Reporter
c	37	1116	89.0	9071	11	AB057644	AB057644 Shuttle v
	38	1116	89.0	12168	11	AF405696	AF405696 Reporter
	39	1116	89.0	12248	11	AF405697	AF405697 Reporter
c	40	1116	89.0	14928	11	AY230218	AY230218 Expressio
c	41	1115.4	88.9	4907	11	SYNPBR328V	L08858 pBR328 clon
c	42	1115.4	88.9	5996	11	SYNPBR325V	L08855 pBR325 clon
	43	1115	88.9	3475	11	D88215	D88215 Cloning vec
	44	1114.4	88.9	4589	11	SYN261CAT	M29363 Plasmid pFL
	45	1114.4	88.9	5838	6	AR072535	AR072535 Sequence

ALIGNMENTS

RESULT 1
AR477289/c
LOCUS AR477289 3400 bp DNA linear PAT 14-MAY-2004
DEFINITION Sequence 1 from patent US 6696278.
ACCESSION AR477289
VERSION AR477289.1 GI:47234624
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 3400)
AUTHORS Carstens,C.-P.
TITLE Method for transfer of DNA segments
JOURNAL Patent: US 6696278-A 1 24-FEB-2004;
Stratagene; La Jolla, CA
FEATURES Location/Qualifiers
source 1. .3400
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN

Query Match 91.5%; Score 1147.2; DB 6; Length 3400;
Best Local Similarity 98.5%; Pred. No. 6.6e-309;
Matches 1158; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3053 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 2994

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2993 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2934

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2933 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2874

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2873 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 2814
```

```
Qy       290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2813 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2754

Qy       350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2753 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 2694

Qy       410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2693 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2634

Qy       470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2633 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2574

Qy       530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2573 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2514

Qy       590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2513 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2454

Qy       650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2453 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2394

Qy       710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2393 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 2334

Qy       770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2333 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2274

Qy       830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2273 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2214

Qy       890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2213 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2154

Qy       950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2153 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 2094

Qy      1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2093 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2034

Qy      1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2033 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1974

Qy      1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1973 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1914

Qy      1190 GAAAAGGATCCGTTTTCGCATTTATCGTGAAACGCT 1225
            |||||||||     |    ||   |  ||| |  ||
Db      1913 GAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT 1878
```

```
RESULT 2
AX554413/c
LOCUS       AX554413               6122 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 1 from Patent WO0246436.
ACCESSION   AX554413
VERSION     AX554413.1  GI:25898196
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Bowdish,K.S., Barbas-Frederickson,S., Wild,M. and Mcwhirter,J.
  TITLE     Novel plasmid vectors
  JOURNAL   Patent: WO 0246436-A 1 13-JUN-2002;
            ALEXION PHARMACEUTICALS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .6122
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid"
ORIGIN

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6122;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-307;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy       1190 GAAAAG 1195
              ||||||
Db        801 GAAAAG 796


RESULT 3
AX554422/c
LOCUS       AX554422               6122 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 10 from Patent WO0246436.
ACCESSION   AX554422
VERSION     AX554422.1  GI:25898199
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Bowdish,K.S., Barbas-Frederickson,S., Wild,M. and Mcwhirter,J.
  TITLE     Novel plasmid vectors
  JOURNAL   Patent: WO 0246436-A 10 13-JUN-2002;
            ALEXION PHARMACEUTICALS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .6122
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid"
ORIGIN

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6122;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-307;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
               |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy         110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy         170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy         230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy         290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy         350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy         410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy         470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy         530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy         590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy         650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy         710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy         770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy         830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy         890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy         950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy        1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy        1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy        1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy        1190 GAAAAG 1195
               ||||||
Db         801 GAAAAG 796


RESULT 4
AX642149/c
LOCUS       AX642149               6122 bp    DNA     linear   PAT 21-FEB-2003
DEFINITION  Sequence 1 from Patent WO0246435.
ACCESSION   AX642149
VERSION     AX642149.1  GI:28474637
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
```

```
  AUTHORS   Bowdish,K.S., Barbas-Frederickson,S., Lin,Y.C., Renshaw,M., Wild,M.
            and Mcwhirter,J.
  TITLE     Engineered plasmids and their use for in situ production of genes
  JOURNAL   Patent: WO 0246435-A 1 13-JUN-2002;
            ALEXION PHARMACEUTICALS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .6122
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="vector"
ORIGIN

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6122;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-307;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
               |||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy         110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy         170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy         230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy         290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy         350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy         410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy         470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy         530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy         590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy         650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy         710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy         770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy         830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy         890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy         950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy        1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy        1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy        1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy        1190 GAAAAG 1195
               ||||||
Db         801 GAAAAG 796


RESULT 5
AX554420/c
LOCUS       AX554420               6125 bp    DNA     linear   PAT 27-NOV-2002
DEFINITION  Sequence 8 from Patent WO0246436.
ACCESSION   AX554420
VERSION     AX554420.1  GI:25898198
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Bowdish,K.S., Barbas-Frederickson,S., Wild,M. and Mcwhirter,J.
  TITLE     Novel plasmid vectors
  JOURNAL   Patent: WO 0246436-A 8 13-JUN-2002;
            ALEXION PHARMACEUTICALS, INC. (US)
FEATURES             Location/Qualifiers
     source          1. .6125
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="plasmid"
ORIGIN

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6125;
  Best Local Similarity   99.8%;  Pred. No. 1.1e-307;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
               |||| ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy         110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy         170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy         230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy         290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy         350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
```

```
Db       1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy       1190 GAAAAG 1195
              ||||||
Db        801 GAAAAG 796
```

```
RESULT 6
CVU46018/c
LOCUS       CVU46018               3399 bp    DNA     linear   SYN 27-OCT-1999
DEFINITION  Cloning vector pCRSCRIPT Cam, complete sequence.
ACCESSION   U46018
VERSION     U46018.1  GI:1184322
KEYWORDS    .
SOURCE      Cloning vector pCRSCRIPT Cam
  ORGANISM  Cloning vector pCRSCRIPT Cam
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 3399)
  AUTHORS   Marsh,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1996) Sam Marsh, Marketing Analysis, Stratagene,
            11011 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES             Location/Qualifiers
     source          1. .3399
                     /organism="Cloning vector pCRSCRIPT Cam"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:45534"
ORIGIN
```

Query Match 90.5%; Score 1135.2; DB 11; Length 3399;
Best Local Similarity 98.4%; Pred. No. 1.5e-305;
Matches 1157; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```
Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3052 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 2993

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2992 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2933

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2932 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2873

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2872 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 2813

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2812 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2753

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2752 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 2693

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2692 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2633

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2632 AC-GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2574

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2573 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2514

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2513 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2454

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2453 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2394

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2393 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 2334

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2333 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2274

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2273 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2214

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2213 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2154
```

```
Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2153 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 2094

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2093 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2034

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1974

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1914

Qy       1190 GAAAAGGATCCGTTTTCGCATTTATCGTGAAACGCT 1225
              |||||||||     |    ||   |  ||| |  ||
Db       1913 GAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT 1878
```

RESULT 7
CS119871/c
LOCUS CS119871 4100 bp DNA linear PAT 08-JUL-2005
DEFINITION Sequence 133 from Patent WO2005056782.
ACCESSION CS119871
VERSION CS119871.1 GI:70667769
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Amin,N.S., Boston,M.G., Bott,R.R., Cervin,M.A., Concar,E.M., Gustwiller,M.E., Jones,B.E., Liebeton,K., Miracle,G.S., Oh,H., Poulose,A.J., Ramer,S.W., Scheibel,J.J., Weyler,W. and Whited,G.M.
TITLE Perhydrolase
JOURNAL Patent: WO 2005056782-A 133 23-JUN-2005; GENENCOR INTERNATIONAL, INC. (US); THE PROCTER & GAMBLE COMPANY (US)
FEATURES Location/Qualifiers
source 1. .4100
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="pAH502R plasmid"
ORIGIN

Query Match 90.5%; Score 1135.2; DB 6; Length 4100;
Best Local Similarity 98.4%; Pred. No. 1.5e-305;
Matches 1157; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```
Qy          50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
               |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        3753 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 3694

Qy         110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3693 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 3634

Qy         170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3633 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 3574

Qy         230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3573 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 3514

Qy         290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3513 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 3454

Qy         350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        3453 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 3394

Qy         410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3393 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 3334

Qy         470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
               || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3333 AC-GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 3275

Qy         530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3274 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 3215

Qy         590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3214 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 3155

Qy         650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3154 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 3095

Qy         710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3094 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 3035

Qy         770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3034 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2975

Qy         830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2974 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2915

Qy         890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2914 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2855

Qy         950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2854 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 2795

Qy        1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2794 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2735

Qy        1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2734 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 2675

Qy        1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2674 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 2615

Qy        1190 GAAAAGGATCCGTTTTCGCATTTATCGTGAAACGCT 1225
               |||||||||     |    ||   |  ||| |  ||
Db        2614 GAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT 2579
```

RESULT 8
AY781408/c
LOCUS AY781408 1808 bp DNA linear SYN 28-FEB-2005
DEFINITION Synthetic construct transposon mini-Mu transposon TnCR1, complete sequence.
ACCESSION AY781408
VERSION AY781408.1 GI:60172112
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 1808)
AUTHORS Zhang,C., Kitsberg,D., Chy,H., Zhou,Q. and Morrison,J.R.
TITLE Transposon-mediated generation of targeting vectors for the

```
          production of gene knockouts
  JOURNAL   (er) Nucleic Acids Res. 33 (3), E24 (2005) In press
    PUBMED  15699181
REFERENCE   2  (bases 1 to 1808)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1808
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1808
                     /transposon="mini-Mu transposon TnCR1"
     CDS             825. .1484
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAX14475.1"
                     /db_xref="GI:60172113"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             90.3%;  Score 1131.8;  DB 11;  Length 1808;
  Best Local Similarity   98.0%;  Pred. No. 1.3e-304;
  Matches 1169;  Conservative    0;  Mismatches   17;  Indels    7;  Gaps    2;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACGGATCC 60
              |||||||     | ||||||||||||||||||||||||||||||||||||||||||
Db       1807 GATCTGAAGCGGCGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACGGCCGG 1748

Qy         61 TAT------CGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCAT 114
                    |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1747 CCGCGGCCGCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCAT 1688

Qy        115 TTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGA 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1687 TTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGA 1628

Qy        175 CATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGA 234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1627 CATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGA 1568

Qy        235 ATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGG 294
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1567 ATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGC-ACCAGGCGTTTAAGG 1509

Qy        295 GCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTG 354
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1508 GCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTG 1449

Qy        355 TAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAAT 414
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1448 TAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAAT 1389

Qy        415 CGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGG 474
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 CGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGG 1329

Qy        475 GGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 GGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG 1269

Qy        535 ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTC 594
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTC 1209

Qy        595 ACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTA 654
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 ACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTA 1149

Qy        655 TTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTG 714
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1148 TTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTG 1089

Qy        715 AACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGC 774
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1088 AACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGC 1029

Qy        775 ATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTT 834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1028 ATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTT 969

Qy        835 TACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGC 894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 TACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGC 909

Qy        895 AACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGT 954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 AACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGT 849

Qy        955 ATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTC 1014
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        848 ATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTC 789

Qy       1015 AAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTG 1074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        788 AAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTG 729

Qy       1075 CCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGAC 1134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 CCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGAC 669

Qy       1135 ACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGG 1187
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        668 ACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGG 616


RESULT 9
AF153422
LOCUS       AF153422                3417 bp    DNA     circular SYN 02-APR-2000
DEFINITION  Cloning vector pTG8, complete sequence.
ACCESSION   AF153422
VERSION     AF153422.1  GI:7384995
KEYWORDS    .
SOURCE      Cloning vector pTG8
  ORGANISM  Cloning vector pTG8
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 3417)
  AUTHORS   Yang,Y. and Spector,A.
  TITLE     Improved cloning vectors for transgene construction
  JOURNAL   BioTechniques 22 (6), 1032-1034 (1997)
    PUBMED  9187746
REFERENCE   2  (bases 1 to 3417)
  AUTHORS   Yang,Y. and Spector,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-1999) Institute of Molecular Biology, University
            of Hong Kong, 8 Sassoon, Pokfulam, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .3417
                     /organism="Cloning vector pTG8"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:119062"
                     /note="derived from Strategene vector pBS KS(+) by
                     insertion of rare FseI and PacI restriction sites flanking
                     the multiple cloning site"
     CDS             complement(241. .834)
                     /codon_start=1
                     /transl_table=11
                     /product="lacZ"
                     /protein_id="AAF61634.1"
```

```
                     /db_xref="GI:7384996"
                     /translation="MTMITPSAQAGQLTLTKGNKSWVPGPPSRSTVSISLISNSCSPG
                     DPLVLERPPPRWSSNSPYSESYYLINARSLAVVLQRRDWENPGVTQLNRLAAHPPFAS
                     WRNSEEARTDRPSQQLRSLNGEWDAPCSGALSAAGVVVTRSVTATLASALAPAPFAFF
                     PSFLATFAGFPRQALNRGLPLGFRFSALRHLDPKKLD"
     CDS             complement(2316. .2867)
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAF61635.1"
                     /db_xref="GI:7384997"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPELRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPFSPWANIIRKATRC"
ORIGIN

  Query Match             90.2%;  Score 1130.8;  DB 11;  Length 3417;
  Best Local Similarity   99.7%;  Pred. No. 2.5e-304;
  Matches 1143;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1940 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1999

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2000 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2059

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2060 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2119

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2120 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 2179

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2180 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2239

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2240 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 2299

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2300 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2359

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2360 AC-GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2418

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2419 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2478

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2479 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2538

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2539 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2598

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2599 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 2658

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2659 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2718

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2719 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2778

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2779 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2838

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2839 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 2898

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2899 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2958

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2959 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 3018

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3019 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 3078

Qy       1190 GAAAAG 1195
              ||||||
Db       3079 GAAAAG 3084


RESULT 10
ACCTN2670L/c
LOCUS       ACCTN2670L              1851 bp    DNA     linear   BCT 01-MAR-1995
DEFINITION  Acinetobacter calcoaceticus Tn2670-like transposon encoding the CAT
            structural gene, complete cds.
ACCESSION   M37690
VERSION     M37690.1  GI:141788
KEYWORDS    CAT gene; Tn2670-like transposon.
SOURCE      Acinetobacter calcoaceticus
  ORGANISM  Acinetobacter calcoaceticus
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Moraxellaceae; Acinetobacter.
REFERENCE   1  (bases 1 to 1851)
  AUTHORS   Elisha,B.G. and Steyn,L.M.
  TITLE     Identification of an Acinetobacter baumannii gene region with
            sequence and organizational similarity to Tn2670
  JOURNAL   Plasmid 25 (2), 96-104 (1991)
   PUBMED   1650008
COMMENT     Original source text: Acinetobacter calcoaceticus anitratus DNA,
            clone pGSH201.
            Draft entry and computer-readable sequence for [Unpublished (1990)]
            kindly submitted
            by L.M.Steyn, 10-AUG-1990.
               Univ of Cape Town
               Dept Med Microbiol
               Med Sch
               Observatory
               Cape Town  7925 South Africa.
FEATURES             Location/Qualifiers
     source          1. .1851
                     /organism="Acinetobacter calcoaceticus"
                     /mol_type="genomic DNA"
                     /isolate="SAK"
                     /sub_species="anitratus"
                     /db_xref="taxon:471"
                     /clone="pGSH201"
     gene            405. .1064
                     /gene="CAT"
     CDS             405. .1064
                     /gene="CAT"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA62571.1"
                     /db_xref="GI:141789"
```

```
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             90.1%;  Score 1129.6;  DB 1;  Length 1851;
  Best Local Similarity   98.8%;  Pred. No. 5.4e-304;
  Matches 1138;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1333 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1274

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1273 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1214

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1154

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1094

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1034

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 974

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 914

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 854

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 794

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        793 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 734

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 674

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        673 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 614

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 554

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 494

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 434

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        433 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 374

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 314

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 254

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 194

Qy       1190 GAAAAGGATCCG 1201
              ||| |   ||
Db        193 CAAAGTGCGTCG 182


RESULT 11
ACCACEAA/c
LOCUS       ACCACEAA                1874 bp    DNA     linear   BCT 24-APR-1993
DEFINITION  A.baumannii chloramphenicol acetyltransferase (cat) gene, complete
            cds.
ACCESSION   M62822
VERSION     M62822.1  GI:141744
KEYWORDS    chloramphenicol acetyltransferase.
SOURCE      Acinetobacter baumannii
  ORGANISM  Acinetobacter baumannii
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Moraxellaceae; Acinetobacter.
REFERENCE   1  (bases 1 to 1874)
  AUTHORS   Elisha,B.G. and Steyn,L.M.
  TITLE     Identification of an Acinetobacter baumannii gene region with
            sequence and organizational similarity to Tn2670
  JOURNAL   Plasmid (1991) In press
COMMENT     Original source text: A.baumannii DNA.
FEATURES             Location/Qualifiers
     source          1. .1874
                     /organism="Acinetobacter baumannii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:470"
     gene            405. .1064
                     /gene="CAT"
     CDS             405. .1064
                     /gene="CAT"
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAA21891.1"
                     /db_xref="GI:141745"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             90.1%;  Score 1129.6;  DB 1;  Length 1874;
  Best Local Similarity   98.8%;  Pred. No. 5.4e-304;
  Matches 1138;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1333 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1274

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1273 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1214

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1213 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1154
```

```
Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1153 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1094

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1093 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1034

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1033 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 974

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        973 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 914

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        913 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 854

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        853 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 794

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        793 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 734

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 674

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        673 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 614

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        613 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 554

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        553 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 494

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 434

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 374

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         373 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 314

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         313 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 254

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         253 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 194

Qy       1190 GAAAAGGATCCG 1201
              |||  |   ||
Db         193 CAAAGTGCGTCG 182


RESULT 12
AY608912
LOCUS       AY608912               6464 bp    DNA     circular BCT 01-JUN-2004
```

```
DEFINITION  Escherichia coli plasmid pFL129, complete sequence.
ACCESSION   AY608912
VERSION     AY608912.1  GI:47717944
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 6464)
  AUTHORS   Wild,J., Czyz,A., Rakowski,S.A. and Filutowicz,M.
  TITLE     Mobilizable gamma ori plasmid pFL129 sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6464)
  AUTHORS   Rakowski,S.A., Wild,J. and Filutowicz,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2004) Bacteriology and Oncology, University of
            Wisconsin - Madison, 420 Henry Mall - Room 151, Madison, WI 53706,
            USA
FEATURES             Location/Qualifiers
     source          1. .6464
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
                     /plasmid="pFL129"
     rep_origin      1. .383
                     /note="gamma origin of plasmid R6K"
     enhancer        1. .106
                     /note="gamma ori enhancer"
     misc_feature    107. .198
                     /note="A+T-rich region"
     protein_bind    199. .352
                     /note="contains 22 bp direct repeat"
                     /bound_moiety="Rep protein pi"
     gene            431. .1348
                     /gene="pir"
     CDS             431. .1348
                     /gene="pir"
                     /note="replication initiator (Rep) for R6K"
                     /codon_start=1
                     /transl_table=11
                     /product="pi"
                     /protein_id="AAT37965.1"
                     /db_xref="GI:47717946"
                     /translation="MRLKVMMDVNKKTKIRHRNELNHTLAQLPLPAKRVMYMALAPID
                     SKEPLERGRVFKIRAEDLAALAKITPSLAYRQLKEGGKLLGASKISLRGDDIIALAKE
                     LNLPFTAKNSPEELDLNIIEWIAYSNDEGYLSLKFTRTIEPYISSLIGKKNKFTTQLL
                     TASLRLSSQYSSSLYQLIRKHYSNFKKKNYFIISVDELKEELIAYTFDKDGNIEYKYP
                     DFPIFKRDVLNKAIAEIKKKTEISFVGFTVHEKEGRKISKLKFEFVVDEDEFSGDKDD
                     EAFFMNLSEADAAFLKVFDETVPPKKAKG"
     variation       750
                     /gene="pir"
                     /note="compared to plasmid pFL130; results in
                     phenylalanine to serine substitution; predicted to produce
                     a copy-up phenotype"
                     /replace="c"
     gene            complement(2136. .2795)
                     /gene="cat"
     CDS             complement(2136. .2795)
                     /gene="cat"
                     /note="chloramphenicol resistance marker"
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAT37967.1"
                     /db_xref="GI:47717948"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
     oriT            3692. .3803
                     /note="from plasmid RK2 (RP4)"
     gene            complement(4180. .5454)
```

```
                     /gene="tetA"
     CDS             complement(4180. .5454)
                     /gene="tetA"
                     /note="tetracycline resistance protein"
                     /codon_start=1
                     /transl_table=11
                     /product="TetA"
                     /protein_id="AAT37966.1"
                     /db_xref="GI:47717947"
                     /translation="MSTNLSVIKNPRVQSDQRRLVRRPDVKPNIPLIVILSTVALDAV
                     GIGLIMPVLPGLLRDLVHSNDVTAHYGILLALYALVQFACAPVLGALSDRFGRRPILL
                     VSLAGATVDYAIMATAPFLWVLYIGRIVAGITGATGAVAGAYIADITDGDERARHFGF
                     MSACFGFGMVAGPVLGGLMGGFSPHAPFFAAAALNGLNFLTGCFLLPESHKGERRPLR
                     REALNPLASFRWARGMTVVAALMAVFFIMQLVGQVPAALWVIFGEDRFHWDATTIGIS
                     LAAFGILHSLAQAMITGPVAARLGERRALMLGMIADGTGYILLAFATRGWMAFPIMVL
                     LASGGIGMPALQAMLSRQVDEERQGQLQGSLAALTSLTSIVGPLLFTAIYAASITTWN
                     GWAWIAGAALYLLCLPALRRGLWSGAGQRADR"
     gene            5458. .6135
                     /gene="tetR"
     CDS             5458. .6135
                     /gene="tetR"
                     /note="tetracycline resistance repressor protein"
                     /codon_start=1
                     /transl_table=11
                     /product="TetR"
                     /protein_id="AAT37964.1"
                     /db_xref="GI:47717945"
                     /translation="MFISDKVSSMTKLQPNTVIRAALDLLNEVGVDGLTTRKLAERLG
                     VQQPALYWHFRNKRALLDALAEAMLAENHTHSVPRADDDWRSFLIGNARSFRQALLAY
                     RDGARIHAGTRPGAPQMETADAQLRFLCEAGFSAGDAVNALMTISYFTVGAVLEEQAG
                     DSDAGERGGTVEQAPLSPLLRAAIDAFDEAGPDAAFEQGLAVIVDGLAKRRLVVRNVE
                     GPRKGDD"
ORIGIN

  Query Match             90.1%;  Score 1129.6;  DB 1;  Length 6464;
  Best Local Similarity   98.8%;  Pred. No. 5.6e-304;
  Matches 1138;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1867 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1926

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1927 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1986

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1987 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2046

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2047 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 2106

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2107 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2166

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2167 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 2226

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2227 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2286

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2287 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2346

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2347 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2406

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2407 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2466

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2467 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2526

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       2527 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 2586

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2587 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2646

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2647 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2706

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2707 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2766

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2767 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 2826

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2827 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2886

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2887 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 2946

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2947 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 3006

Qy       1190 GAAAAGGATCCG 1201
              ||| |   ||
Db       3007 CAAAGTGCGTCG 3018


RESULT 13
AY303237
LOCUS       AY303237               11925 bp    DNA     circular SYN 02-APR-2004
DEFINITION  Shuttle vector pLEIItd+KR''deltaORF, complete sequence.
ACCESSION   AY303237
VERSION     AY303237.1  GI:32185266
KEYWORDS    .
SOURCE      Shuttle vector pLEIItd+KR''deltaORF
  ORGANISM  Shuttle vector pLEIItd+KR''deltaORF
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 11925)
  AUTHORS   Staddon,J.H., Bryan,E.M., Manias,D.A. and Dunny,G.M.
  TITLE     Conserved Target for Group II Intron Insertion in Relaxase Genes of
            Conjugative Elements of Gram-Positive Bacteria
  JOURNAL   J. Bacteriol. 186 (8), 2393-2401 (2004)
   PUBMED   15060042
REFERENCE   2  (bases 1 to 11925)
  AUTHORS   Staddon,J.H.
  TITLE     Ll.ltrB group II intron donor plasmid with a large deletion in the
            intron encoded protein ORF, ltrA
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 11925)
  AUTHORS   Staddon,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2003) Microbiology, University of Minnesota, MMC
            196, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
FEATURES             Location/Qualifiers
```

```
source          1. .11925
                /organism="Shuttle vector pLEIItd+KR''deltaORF"
                /mol_type="genomic DNA"
                /db_xref="taxon:234876"
exon            1623. .1802
                /note="ltrBE1; partial exon"
                /number=1
intron          1803. .4225
                /note="Ll.ltrBdeltaORF group II intron with APHIII and
                tddelta(1-3)"
CDS             2376. .2585
                /note="maturase with internal deletion"
                /codon_start=1
                /transl_table=11
                /product="LtrAdeltaORF"
                /protein_id="AAP73797.1"
                /db_xref="GI:32185267"
                /translation="MKPTMAILERISKNSQENIDEVFTRLYRYLLRPDIYYVATRWEM
                AMIAKQRKTLVVCFHCHRHVIHKHK"
exon            2594. .2605
                /note="derived from T4 phage thymidylate synthase found in
                GenBank Accession Number M12742"
                /number=1
intron          2606. .2998
                /note="tddelta(1-3) group I intron"
exon            2999. .3010
                /note="derived from T4 phage thymidylate synthase found in
                GenBank Accession Number M12742; partial exon"
                /number=2
CDS             3255. .4049
                /note="3',5'aminoglycoside phosphotransferase; kanamycin
                resistance; from enterococcal plasmid pJH1 found in
                GenBank Accession Number V01547"
                /codon_start=1
                /transl_table=11
                /product="APH-III"
                /protein_id="AAP73802.1"
                /db_xref="GI:32185272"
                /translation="MAKMRISPELKKLIEKYRCVKDTEGMSPAKVYKLVGENENLYLK
                MTDSRYKGTTYDVEREKDMMLWLEGKLPVPKVLHFERHDGWSNLLMSEADGVLCSEEY
                EDEQSPEKIIELYAECIRLFHSIDISDCPYTNSLDSRLAELDYLLNNDLADVDCENWE
                EDTPFKDPRELYDFLKTEKPEEELVFSHGDLGDSNIFVKDGKVSGFIDLGRSGRADKW
                YDIAFCVRSIREDIGEEQYVELFFDLLGIKPDWEKIKYYILLDELF"
exon            4226. .4316
                /note="ltrBE2; partial exon"
                /number=2
CDS             complement(5166. .6656)
                /note="pIL253 replication in gram positive bacteria;
                derived from GenBank Accession Number AF041239"
                /codon_start=1
                /transl_table=11
                /product="replication protein E"
                /protein_id="AAP73798.1"
                /db_xref="GI:32185268"
                /translation="MNIPFVVETVLHDGLLKYKFKNSKIRSITTKPGKSKGAIFAYRS
                KSSMIGGRGVVLTSEEAIQENQDTFTHWTPNVYRYGTYADENRSYTKGHSENNLRQIN
                TFFIDFDIHTAKETISASDILTTAIDLGFMPTMIIKSDKGYQAYFVLETPVYVTSKSE
                FKSVKAAKIISQNIREYFGKSLPVDLTCNHFGIARIPRTDNVEFFDPNYRYSFKEWQD
                WSFKQTDNKGFTRSSLTVLSGTEGKKQVDEPWFNLLLHETKFSGEKGLIGRNNVMFTL
                SLAYFSSGYSIETCEYNMFEFNNRLDQPLEEKEVIKIVRSAYSENYQGANREYITILC
                KAWVSSDLTSKDLFVRQGWFKFKKKRSERQRVHLSEWKEDLMAYISEKSDVYKPYLVT
                TKKEIREVLGIPERTLDKLLKVLKANQEIFFKIKPGRNGGIQLASVKSLLLSIIKVKK
                EEKESYIKALTNSFDLEHTFIQETLNKLAERPKTDTQLDLFSYDTG"
CDS             complement(6699. .6998)
                /note="pIL253 replication in gram positive bacteria;
                derived from GenBank Accession Number AF041239"
                /codon_start=1
                /transl_table=11
                /product="replication protein D"
                /protein_id="AAP73799.1"
                /db_xref="GI:32185269"
                /translation="MNHDECKTYIKNSLLEIRKLANIYTLETFKKELEKRNIYLETKS
                DKYFSSEGEDYIYKLIENNKIIYSISGKKLTYKGKKSFSKHAILKQLNEKANQVN"
CDS             8339. .8989
                /note="functions in gram positive bacteria; derived from
                pC194 found in GenBank Accession Number V01277"
                /codon_start=1
                /transl_table=11
                /product="chloramphenicol acetyltransferase"
                /protein_id="AAP73800.1"
                /db_xref="GI:32185270"
                /translation="MNFNKIDLDNWKRKEIFNHYLNQQTTFSITTEIDISVLYRNIKQ
                EGYKFYPAFIFLVTRVINSNTAFRTGYNSDGELGYWDKLEPLYTIFDGVSKTFSGIWT
                PVKNDFKEFYDLYLSDVEKYNGSGKLFPKTPIPENAFSLSIIPWTSFTGFNLNINNNS
                NYLLPIITAGKFINKGNSIYLPLSLQVHHSVCDGYHAGLFMNSIQELSDRPNDWLL"
CDS             complement(10617. .11276)
                /note="functions in gram negative bacteria; derived from
                plasmid pACYC184 found in GenBank Accession Number X06403"
                /codon_start=1
                /transl_table=11
                /product="chloramphenicol acetyltransferase"
                /protein_id="AAP73801.1"
                /db_xref="GI:32185271"
                /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                A"
ORIGIN

  Query Match           90.1%;  Score 1129.6;  DB 11;  Length 11925;
  Best Local Similarity  98.8%;  Pred. No. 5.7e-304;
  Matches 1138;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      10348 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 10407

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10408 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 10467

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10468 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 10527

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10528 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 10587

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10588 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 10647

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10648 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 10707

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10708 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 10767

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10768 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 10827

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10828 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 10887

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10888 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 10947

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      10948 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 11007

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      11008 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 11067

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11068 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 11127

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11128 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 11187

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11188 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 11247

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11248 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 11307

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11308 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 11367

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11368 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 11427

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11428 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 11487

Qy       1190 GAAAAGGATCCG 1201
              |||  |   ||
Db      11488 CAAAGTGCGTCG 11499
```

```
RESULT 14
AY303236
LOCUS       AY303236               13515 bp    DNA     circular SYN 02-APR-2004
DEFINITION  Shuttle vector pLEIItd+KR'', complete sequence.
ACCESSION   AY303236
VERSION     AY303236.1  GI:32140767
KEYWORDS    .
SOURCE      Shuttle vector pLEIItd+KR''
  ORGANISM  Shuttle vector pLEIItd+KR''
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 13515)
  AUTHORS   Cousineau,B., Lawrence,S., Smith,D. and Belfort,M.
  TITLE     Retrotransposition of a bacterial group II intron
  JOURNAL   Nature 404 (6781), 1018-1021 (2000)
   PUBMED   10801134
REFERENCE   2  (bases 1 to 13515)
  AUTHORS   Staddon,J.H., Bryan,E.M., Manias,D.A. and Dunny,G.M.
  TITLE     Conserved Target for Group II Intron Insertion in Relaxase Genes of
            Conjugative Elements of Gram-Positive Bacteria
  JOURNAL   J. Bacteriol. 186 (8), 2393-2401 (2004)
   PUBMED   15060042
REFERENCE   3  (bases 1 to 13515)
  AUTHORS   Staddon,J.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2003) Microbiology, University of Minnesota, MMC
            196, 420 Delaware Street S.E., Minneapolis, MN 55455, USA
FEATURES             Location/Qualifiers
     source          1. .13515
                     /organism="Shuttle vector pLEIItd+KR''"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:233895"
     exon            1623. .1802
                     /note="ltrBE1; partial exon"
                     /number=1
     intron          1803. .5815
                     /note="Ll.ltrB group II intron with APHIII and td
                     delta(1-3)"
     CDS             2376. .4175
                     /note="maturase"
                     /codon_start=1
                     /transl_table=11
                     /product="LtrA"
                     /protein_id="AAP70045.1"
                     /db_xref="GI:32140772"
                     /translation="MKPTMAILERISKNSQENIDEVFTRLYRYLLRPDIYYVAYQNLY
                     SNKGASTKGILDDTADGFSEEKIKKIIQSLKDGTYYPQPVRRMYIAKKNSKKMRPLGI
                     PTFTDKLIQEAVRIILESIYEPVFEDVSHGFRPQRSCHTALKTIKREFGGARWFVEGD
                     IKGCFDNIDHVTLIGLINLKIKDMKMSQLIYKFLKAGYLENWQYHKTYSGTPQGGILS
                     PLLANIYLHELDKFVLQLKMKFDRESPERITPEYRELHNEIKRISHRLKKLEGEEKAK
                     VLLEYQEKRKRLPTLPCTSQTNKVLKYVRYADDFIISVKGSKEDCQWIKEQLKLFIHN
                     KLKMELSEEKTLITHSSQPARFLGYDIRVRRSGTIKRSGKVKKRTLNGSVELLIPLQD
                     KIRQFIFDKKIAIQKKDSSWFPVHRKYLIRSTDLEIITIYNSELRGICNYYGLASNFN
                     QLNYFAYLMEYSCLKTIASKHKGTLSKTISMFKDGSGSWGIPYEIKQGKQRRYFANFS
                     ECKSPYQFTDEISQAPVLYGYARNTLENRLKAKCCELCGTSDENTSYEIHHVNKVKNL
                     KGKEKWEMAMIAKQRKTLVVCFHCHRHVIHKHK"
     exon            4184. .4195
                     /note="derived from T4 phage thymidylate synthase of
                     GenBank Accession Number M12742"
                     /number=1
     intron          4196. .4588
                     /note="td delta(1-3) group I intron"
     exon            4589. .4600
                     /note="derived from T4 phage thymidylate synthase of
                     GenBank Accession Number M12742; partial exon"
                     /number=2
     CDS             4845. .5639
                     /note="3',5''aminoglycoside phosphotransferase; kanamycin
                     resistance; from enterococcal plasmid pJH1 of GenBank
                     Accession Number V01547"
                     /codon_start=1
                     /transl_table=11
                     /product="APH-III"
                     /protein_id="AAP70046.1"
                     /db_xref="GI:32140773"
                     /translation="MAKMRISPELKKLIEKYRCVKDTEGMSPAKVYKLVGENENLYLK
                     MTDSRYKGTTYDVEREKDMMLWLEGKLPVPKVLHFERHDGWSNLLMSEADGVLCSEEY
                     EDEQSPEKIIELYAECIRLFHSIDISDCPYTNSLDSRLAELDYLLNNDLADVDCENWE
                     EDTPFKDPRELYDFLKTEKPEEELVFSHGDLGDSNIFVKDGKVSGFIDLGRSGRADKW
                     YDIAFCVRSIREDIGEEQYVELFFDLLGIKPDWEKIKYYILLDELF"
     exon            5816. .5906
                     /note="ltrBE2; partial exon"
                     /number=2
     CDS             complement(6756. .8246)
                     /note="derived from GenBank Accession Number AF041239"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein E"
                     /protein_id="AAP70043.1"
                     /db_xref="GI:32140770"
                     /translation="MNIPFVVETVLHDGLLKYKFKNSKIRSITTKPGKSKGAIFAYRS
                     KSSMIGGRGVVLTSEEAIQENQDTFTHWTPNVYRYGTYADENRSYTKGHSENNLRQIN
                     TFFIDFDIHTAKETISASDILTTAIDLGFMPTMIIKSDKGYQAYFVLETPVYVTSKSE
                     FKSVKAAKIISQNIREYFGKSLPVDLTCNHFGIARIPRTDNVEFFDPNYRYSFKEWQD
                     WSFKQTDNKGFTRSSLTVLSGTEGKKQVDEPWFNLLLHETKFSGEKGLIGRNNVMFTL
                     SLAYFSSGYSIETCEYNMFEFNNRLDQPLEEKEVIKIVRSAYSENYQGANREYITILC
                     KAWVSSDLTSKDLFVRQGWFKFKKKRSERQRVHLSEWKEDLMAYISEKSDVYKPYLVT
                     TKKEIREVLGIPERTLDKLLKVLKANQEIFFKIKPGRNGGIQLASVKSLLLSIIKVKK
                     EEKESYIKALTNSFDLEHTFIQETLNKLAERPKTDTQLDLFSYDTG"
     CDS             complement(8289. .8588)
                     /note="pIL253 replication in gram positive bacteria;
                     derived from GenBank Accession Number AF041239"
                     /codon_start=1
                     /transl_table=11
                     /product="replication protein D"
                     /protein_id="AAP70044.1"
                     /db_xref="GI:32140771"
                     /translation="MNHDECKTYIKNSLLEIRKLANIYTLETFKKELEKRNIYLETKS
```

```
                     DKYFSSEGEDYIYKLIENNKIIYSISGKKLTYKGKKSFSKHAILKQLNEKANQVN"
     CDS             9929. .10579
                     /note="functions in gram positive bacteria; derived from
                     pC194 of GenBank Accession Number V01277"
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAP70041.1"
                     /db_xref="GI:32140768"
                     /translation="MNFNKIDLDNWKRKEIFNHYLNQQTTFSITTEIDISVLYRNIKQ
                     EGYKFYPAFIFLVTRVINSNTAFRTGYNSDGELGYWDKLEPLYTIFDGVSKTFSGIWT
                     PVKNDFKEFYDLYLSDVEKYNGSGKLFPKTPIPENAFSLSIIPWTSFTGFNLNINNNS
                     NYLLPIITAGKFINKGNSIYLPLSLQVHHSVCDGYHAGLFMNSIQELSDRPNDWLL"
     CDS             complement(12207. .12866)
                     /note="functions in gram negative bacteria; derived from
                     plasmid pACYC184 of GenBank Accession Number X06403"
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAP70042.1"
                     /db_xref="GI:32140769"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match             90.1%;  Score 1129.6;  DB 11;  Length 13515;
  Best Local Similarity   98.8%;  Pred. No. 5.7e-304;
  Matches 1138;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11938 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 11997

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11998 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 12057

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12058 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 12117

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12118 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 12177

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12178 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 12237

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12238 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 12297

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12298 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 12357

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12358 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 12417

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12418 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 12477

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12478 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 12537

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12538 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 12597

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      12598 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 12657

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12658 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 12717

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12718 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 12777

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12778 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 12837

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12838 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 12897

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12898 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 12957

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12958 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 13017

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13018 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 13077

Qy       1190 GAAAAGGATCCG 1201
              |||  |   ||
Db      13078 CAAAGTGCGTCG 13089
```

```
RESULT 15
AY123253/c
LOCUS       AY123253               45325 bp    DNA     linear   BCT 20-DEC-2004
DEFINITION  Klebsiella pneumoniae plasmid pRMH760, partial sequence.
ACCESSION   AY123253 AY242531 AY242532 AY242533
VERSION     AY123253.3  GI:40795472
KEYWORDS    .
SOURCE      Klebsiella pneumoniae
  ORGANISM  Klebsiella pneumoniae
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Klebsiella.
REFERENCE   1  (bases 4901 to 22900)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     In34, a complex In5 family class 1 integron containing orf513 and
            dfrA10
  JOURNAL   Antimicrob. Agents Chemother. 47 (1), 342-349 (2003)
   PUBMED   12499211
REFERENCE   2  (bases 1 to 4900; 29852 to 31416; 33537 to 35101; 43822 to
            45325)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     The IS1111 family members IS4321 and IS5075 have subterminal
            inverted repeats and target the terminal inverted repeats of Tn21
            family transposons
  JOURNAL   J. Bacteriol. 185 (21), 6371-6384 (2003)
   PUBMED   14563872
REFERENCE   3  (bases 1 to 4900)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     The IS1111 family members IS4321 and IS5075 have subterminal
            inverted repeats and target the terminal inverted repeats of Tn21
            family transposons
  JOURNAL   J. Bacteriol. 185 (21), 6371-6384 (2003)
   PUBMED   14563872
REFERENCE   4  (bases 1 to 45286)
```

```
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     Complex multiple antibiotic and mercury resistance region derived
            from the r-det of NR1 (R100)
  JOURNAL   Antimicrob. Agents Chemother. 48 (11), 4250-4255 (2004)
   PUBMED   15504849
REFERENCE   5  (bases 4901 to 29360)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2002) CSIRO Molecular Science, PO Box 184, North
            Ryde, NSW 1670, Australia
REFERENCE   6  (bases 1 to 4900)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2003) CSIRO Molecular Science, PO Box 184, North
            Ryde, NSW 1670, Australia
  REMARK    Sequence update by submitter
REFERENCE   7  (bases 1 to 4900)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) CSIRO Molecular Science, PO Box 184, North
            Ryde, NSW 1670, Australia
REFERENCE   8  (bases 1 to 45325)
  AUTHORS   Partridge,S.R. and Hall,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JAN-2004) Department of Biological Sciences,
            Macquarie University, Sydney, NSW 2109, Australia
  REMARK    Sequence update by submitter
COMMENT     On or before Dec 20, 2004 this sequence version replaced
            gi:33416297, gi:33416299, gi:33416301, gi:28558819.
FEATURES             Location/Qualifiers
     source          1. .45325
                     /organism="Klebsiella pneumoniae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:573"
                     /plasmid="pRMH760"
     misc_feature    join(124. .145,1473. .5119)
                     /note="tnp module; identical to Tn1696 tnp module in
                     Genbank Accession Number U12338.2, with IS4321L in IRtnp"
     repeat_region   join(124. .145,1473. .1488)
                     /note="IRtnp of Tn1696"
                     /rpt_type=inverted
     repeat_region   146. .1472
                     /citation=[2]
                     /insertion_seq="IS4321L"
     repeat_region   149. .160
                     /note="IRr of IS4321"
                     /rpt_type=inverted
     CDS             complement(401. .1405)
                     /codon_start=1
                     /transl_table=11
                     /product="IS4321 transposase"
                     /protein_id="AAO46046.1"
                     /db_xref="GI:28558821"
                     /translation="MENIALIGIDLGKNSFHIHCQDRRGKAVYRKKFTRPKLIEFLAT
                     CPATTIAMEACGGSHFMARKLEELGHFPKLISPQFVRPFVKSNKNDFVDAEAICEAAS
                     RPSMRFVQPRTESQQAMRALHRVRESLVQDKVKTTNQMHAFLLEFGISVPRGAAVISR
                     LSTLLEDNSLPLYLSQLLLKLQQHYHYLVEQIKDLESQLKRKLDEDEIGQRLLSIPCV
                     GTLTASTISTEIGDGKQYASSRDFAAATGLVPRQYSTGGRTTLLGISKRGNKKIRTLL
                     VQCARVFIQKLEHQSGKLADWVRDLLCRKSNFVVTCALANKLARIAWALTARQQTYVA
                     "
     repeat_region   1454. .1465
                     /note="IR1 of IS4321"
                     /rpt_type=inverted
     gene            complement(1484. .4456)
                     /gene="tnpA"
     CDS             complement(1484. .4456)
                     /gene="tnpA"
                     /note="Tn1696 tnpA"
                     /codon_start=1
                     /transl_table=11
                     /product="TnpA transposase"
                     /protein_id="AAO46047.1"
                     /db_xref="GI:28558822"
                     /translation="MPRRSILSATERESLLALPDAKDELIRHYTFNETDLSVIRQRRG
                     AANRLGFAVQLCYLRFPGTFLGVDEPPFPPLLRMVAAQLKMPVESWSEYGQREQTRRE
                     HLVELQTVFGFKPFTMSHYRQAVHTLTELALQTDKGIVLASALVENLRRQSIILPAMN
                     AIERASAEAITRANRRIYAALTDSLLSPHRQRLDELLKRKDGSKVTWLAWLRQSPAKP
                     NSRHMLEHIERLKSWQALDLPAGIERQVHQNRLLKIAREGGQMTPADLAKFEVQRRYA
                     TLVALAIEGMATVTDEIIDLHDRIIGKLFNAAKNKHQQQFQASGKAINDKVRMYGRIG
                     QALIEAKQSGSDPFAAIEAVMPWDTFAASVTEAQTLARPADFDFLHHIGESYATLRRY
                     APQFLGVLKLRAAPAAKGVLDAIDMLRGMNSDSARKVPADAPTAFIKPRWAKLVLTDD
                     GIDRRYYELCALSELKNALRSGDVWVQGSRQFKDFDEYLVPVEKFATLKLASELPLAV
                     ATDCDQYLHDRLELLEAQLATVNRMAAANDLPDAIITTASGLKITPLDAAVPDAAQAM
                     IDQTAMLLPHLKITELLMEVDEWTGFTRHFTHLKTSDTAKDKTLLLTTILADAINLGL
                     TKMAESCPGTTYAKLSWLQAWHIRDETYSTALAELVNAQFRQPFAGNWGDGTTSSSDG
                     QNFRTGSKAESTGHINPKYGSSPGRTFYTHISDQYAPFSAKVVNVGIRDSTYVLDGLL
                     YHESDLRIEEHYTDTAGFTDHVFGLMHLLGFRFAPRIRDLGETKLFIPKGDAAYDALK
                     PMISSDRLNIKQIRAHWDEILRLATSIKQGTVTASLMLRKLGSYPRQNGLAVALRELG
                     RIERTLFILDWLQSVELRRRVHAGLNKGEARNALARAVFFYRLGEIRDRSFEQQRYRA
                     SGLNLVTAAIVLWNTVYLERATSALRGNGTALDDTLLQYLSPLGWEHINLTGDYLWRS
                     SAKVGAGKFRPLRPLPPA"
     gene            complement(4459. .5016)
                     /gene="tnpR"
     CDS             complement(4459. .5016)
                     /gene="tnpR"
                     /note="Tn1696 tnpR"
                     /codon_start=1
                     /transl_table=11
                     /product="TnpR resolvase"
                     /protein_id="AAM89413.2"
                     /db_xref="GI:28558820"
                     /translation="MQGQRIGYVRVSSFDQNPERQLEGVQVARVFTDKASGKDTQRPE
                     LERLLAFVREGDTVVVHSMDRLARNLDDLRRIVQGLTQRGVRMEFVKEGLKFTGEDSP
                     MANLMLSVMGAFAEFERALIRERQREGIVLAKQRGAYRGRKKSLNSEQIAELKRRVAA
                     GDQKTLVARDFGISRETLYQYLRED"
     misc_recomb     5037. .5058
                     /note="Tn1696 resIII"
     misc_recomb     5072. .5105
                     /note="Tn1696 resII"
     repeat_region   5120. .22649
                     /transposon="class I integron In34"
     misc_feature    5120. .6471
                     /note="5'-CS; 5'-conserved segment"
     repeat_region   5120. .5144
                     /note="left-hand inverted repeat of In34; IRi"
                     /rpt_type=inverted
     gene            complement(5322. .6335)
                     /gene="intI1"
     CDS             complement(5322. .6335)
                     /gene="intI1"
                     /codon_start=1
                     /transl_table=11
                     /product="IntI1 integrase"
                     /protein_id="AAM89398.1"
                     /db_xref="GI:24527229"
                     /translation="MKTATAPLPPLRSVKVLDQLRERIRYLHYSLRTEQAYVNWVRAF
                     IRFHGVRHPATLGSSEVEAFLSWLANERKVSVSTHRQALAALLFFYGKVLCTDLPWLQ
                     EIGRPRPSRRLPVVLTPDEVVRILGFLEGEHRLFAQLLYGTGMRISEGLQLRVKDLDF
                     DHGTIIVREGKGSKDRALMLPESLAPSLREQLSRARAWWLKDQAEGRSGVALPDALER
                     KYPRAGHSWPWFWVFAQHTHSTDPRSGVVRRHHMYDQTFQRAFKRAVEQAGITKPATP
                     HTLRHSFATALLRSGYDIRTVQDLLGHSDVSTTMIYTHVLKVGGAGVRSPLDALPPLT
                     SER"
     promoter        6220. .6248
                     /note="promoter Pc; transcribes cassette associated genes;
                     strong version"
     -35_signal      6220. .6225
     -10_signal      6243. .6248
     -10_signal      complement(6362. .6367)
                     /gene="intI1"
     -35_signal      complement(6385. .6390)
                     /gene="intI1"
     misc_recomb     6416. .6480
                     /note="attI1; integron-associated recombination site
                     recognized by IntI1"
     misc_feature    6472. .7062
                     /note="aadB gene cassette"
```

```
     misc_recomb     join(7009. .7062,6472. .6477)
                     /note="aadB 59-be; 59-base element; IntI-type integrase
                     recombination site"
     gene            6481. .7014
                     /gene="aadB"
     CDS             6481. .7014
                     /gene="aadB"
                     /note="confers resistance to gentamicin, kanamycin and
                     tobramycin"
                     /codon_start=1
                     /transl_table=11
                     /product="aminoglycoside (2'') adenylyltransferase
                     AAD(2'')"
                     /protein_id="AAM89399.1"
                     /db_xref="GI:24527230"
                     /translation="MDTTQVTLIHKILAAADERNLPLWIGGGWAIDARLGRVTRKHDD
                     IDLTFPGERRGELEAIVEMLGGRVMEELDYGFLAEIGDELLDCEPAWWADEAYEIAEA
                     PQGSCPEAAEGVIAGRPVRCNSWEAIIWDYFYYADEVPPVDWPTKHIESYRLACTSLG
                     AEKVEVLRAAFRSRYAA"
     misc_feature    7063. .8375
                     /note="3'-CS1; first copy of 3'-conserved segment; bases
                     1-1313 of standard 3'-CS"
     misc_feature    7063. .7452

  Query Match             90.1%;  Score 1129.6;  DB 1;  Length 45325;
  Best Local Similarity   98.8%;  Pred. No. 5.8e-304;
  Matches 1138;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |  ||   |||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      33294 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 33235

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33234 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 33175

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33174 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 33115

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33114 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 33055

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33054 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 32995

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32994 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 32935

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32934 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 32875

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32874 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 32815

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32814 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 32755

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32754 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 32695

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32694 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 32635

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db      32634 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 32575

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32574 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 32515

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32514 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 32455

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32454 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 32395

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32394 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 32335

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32334 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 32275

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32274 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 32215

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      32214 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 32155

Qy       1190 GAAAAGGATCCG 1201
                |||  |   ||
Db      32154 CAAAGTGCGTCG 32143
```

Search completed: January 18, 2006, 04:25:16
Job time : 6465.19 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 779.788 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-2
Perfect score: 1254
Sequence: 1 gatctgattgattgaacgaa.....ttcgttcaatcaatcagatc 1254

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1254	100.0	1254	10	ACF58169	Acf58169 Modified
c 2	1147.2	91.5	3400	7	ADL18593	Adl18593 Plasmid p
c 3	1142.8	91.1	6122	6	ABN84078	Abn84078 Plasmid p
c 4	1142.8	91.1	6122	6	AAL41112	Aal41112 Plasmid p
c 5	1142.8	91.1	6122	6	AAL41115	Aal41115 Plasmid p
c 6	1142.8	91.1	6125	6	AAL41114	Aal41114 Plasmid p
c 7	1135.2	90.5	4100	14	AEA43207	Aea43207 M. smegma
c 8	1130.8	90.2	3817	10	ADD14893	Add14893 Phage dis
c 9	1116	89.0	6607	14	ADZ03868	Adz03868 Plasmid P
c 10	1116	89.0	6608	14	ADZ03870	Adz03870 Plasmid P
c 11	1116	89.0	9808	14	ADZ66224	Adz66224 Construct
c 12	1116	89.0	9808	14	AEB80052	Aeb80052 Vector pA
c 13	1116	89.0	9808	14	AEB72933	Aeb72933 DNA seque
14	1114.4	88.9	5838	2	AAV28398	Aav28398 Plasmid p
15	1114.4	88.9	5838	2	AAZ06434	Aaz06434 Plasmid p
16	1114.4	88.9	5838	2	AAZ22882	Aaz22882 Plasmid p
c 17	1114.4	88.9	7814	6	ABK87209	Abk87209 Synthetic
18	1114.2	88.9	3064	14	ADV09403	Adv09403 Plasmid p
19	1106.6	88.2	5300	14	ADY91370	Ady91370 E. coli e
20	1105.6	88.2	5777	14	ADW46688	Adw46688 Plasmid p
c 21	1074.8	85.7	1853	2	AAQ13156	Aaq13156 Self-clea
c 22	1070.4	85.4	1864	6	AAD35079	Aad35079 GeneHunte
c 23	1052.4	83.9	1906	10	ADF44789	Adf44789 De novo s
c 24	1052.4	83.9	1992	10	ADF44788	Adf44788 De novo s
c 25	1052.4	83.9	2191	10	ADF44787	Adf44787 De novo s
c 26	1052.4	83.9	2714	10	ADF44786	Adf44786 De novo s
c 27	998.2	79.6	9265	6	ABK87210	Abk87210 Synthetic
c 28	982.6	78.4	4626	2	AAT79498	Aat79498 Plasmid p
c 29	982.6	78.4	5227	2	AAT79537	Aat79537 Plasmid p
c 30	982.6	78.4	5362	14	ADV44285	Adv44285 pFosbeta
c 31	982.6	78.4	5362	14	ADV44286	Adv44286 pJunbeta
c 32	982.6	78.4	5563	14	AEB31043	Aeb31043 Plasmid D
c 33	982.6	78.4	5587	14	AEB31042	Aeb31042 Plasmid D
c 34	982.6	78.4	7783	2	AAX26302	Aax26302 Sequence
c 35	982.6	78.4	8623	4	AAF77067	Aaf77067 pCKZEBB v
c 36	960	76.6	9460	5	AAS00466	Aas00466 Plasmid p
37	952.8	76.0	7507	2	AAX01427	Aax01427 Plasmid v
38	952.8	76.0	12495	12	ADO05573	Ado05573 Nucleotid
39	952.8	76.0	14468	8	ACC83500	Acc83500 Bacterial
40	952.8	76.0	16847	12	ADO07464	Ado07464 Japanese
41	952.8	76.0	18563	12	ADO07466	Ado07466 Japanese
42	952.8	76.0	18563	12	ADO07465	Ado07465 Japanese
43	952.8	76.0	18565	12	ADO07467	Ado07467 Japanese
44	952.8	76.0	19038	12	ADO07468	Ado07468 Japanese
45	952.8	76.0	19038	12	ADO07469	Ado07469 Japanese

ALIGNMENTS

RESULT 1
ACF58169
ID ACF58169 standard; DNA; 1254 BP.
XX
AC ACF58169;
XX
DT 15-JAN-2004 (first entry)
XX
DE Modified Cat-Mu(Stop)-transposon.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Claim 6; Page 30; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a

```
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
SQ   Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 1254;  DB 10;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1254;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACGGATCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GATCTGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAACGGATCC 60

Qy         61 TATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCATTTGAGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCATTTGAGA 120

Qy        121 AGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAG 180

Qy        181 CGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATTTCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATTTCT 240

Qy        241 GCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCA 300

Qy        301 ATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTC 360

Qy        361 ATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAG 420

Qy        421 CGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAA 480

Qy        481 GAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGC 540

Qy        541 TGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTA 600

Qy        601 ACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACT 660

Qy        661 CCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACT 720

Qy        721 ATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCAT 780

Qy        781 CAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGT 840

Qy        841 CTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGA 900

Qy        901 CTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCC 960

Qy        961 AGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAA 1020

Qy       1021 TACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATC 1080

Qy       1081 AACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGG 1140

Qy       1141 ATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTCGAAAAGGATCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTCGAAAAGGATCC 1200

Qy       1201 GTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTTCAATCAATCAGATC 1254
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTTCAATCAATCAGATC 1254


RESULT 2
ADL18593/c
ID   ADL18593 standard; DNA; 3400 BP.
XX
AC   ADL18593;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plasmid pBC SK+ nucleotide sequence SEQ ID NO:1.
XX
KW   gene transfer; product vector; purification; detection; plasmid; pBC SK+;
KW   circular; gene; ds.
XX
OS   Synthetic.
XX
PN   WO200268670-A1.
XX
PD   06-SEP-2002.
XX
PF   15-FEB-2002; 2002WO-US004454.
XX
PR   26-FEB-2001; 2001US-00793372.
XX
PA   (STRA-) STRATAGENE.
XX
PI   Carstens C;
XX
DR   WPI; 2002-707008/76.
XX
PT   Transferring a gene to a product vector, useful for generating
PT   recombinant vectors for expression in cell or host, comprises contacting
PT   in vitro a first and second vector and introducing the co-integrate
PT   vector into a prokaryotic cell.
XX
PS   Example 1; SEQ ID NO 1; 50pp; English.
XX
CC   The present invention describes a method for transferring genes to a
CC   product vector comprising contacting in vitro a first and second vector
CC   and introducing the co-integrate vector into a prokaryotic cell to permit
CC   the formation of a product vector. Also described: (1) a pair of vectors
CC   (I) comprising: (a) a first vector comprising a gene or a cloning site
CC   for the insertion of a gene, a gene encoding a first selectable marker, a
CC   double-stranded origin of replication and a site-specific recombination
CC   recognition site, where the gene is interposed between the double-
CC   stranded origin of replication of a rolling circle replicon and the site-
CC   specific recombination recognition site; and (b) a second vector
CC   comprising a negative selectable marker, a double-stranded and double-
CC   stranded origin of replication of a rolling circle replicon, a site-
CC   specific recombination recognition site and a gene encoding a second
```

```
CC   selectable marker, where the gene encoding the negative selectable marker
CC   is interposed between the double-stranded origin of replication of a
CC   rolling circle replicon and the site-specific recognition site, where one
CC   or both of the vectors have no second site-specific recombinase
CC   recognition site between the double-stranded origin of replication and
CC   the site-specific recombination site; (2) a product vector comprising a
CC   gene, a double-stranded origin of replication of a rolling circle
CC   replicon, a site-specific recombination site, a single-stranded origin of
CC   replication and a nucleic aid sequence encoding a second selectable
CC   marker, where the gene is interposed between the double-stranded origin
CC   of replication of a rolling circle replicon and the site-specific
CC   recombination recognition site; and (3) a kit for the transfer of gene to
CC   a product vector comprising (I) and packaging materials. The method is
CC   useful for generating recombinant vectors. These recombinant vectors are
CC   useful in expressing mammalian cell and bacterial hosts, purification of
CC   the native protein by employing specialised purification tags and
CC   detection of interaction with other proteins. The present sequence
CC   represents the nucleotide sequence of plasmid pBC SK+, which is used in
CC   the exemplification of the present invention.
XX
SQ   Sequence 3400 BP; 877 A; 804 C; 863 G; 856 T; 0 U; 0 Other;

  Query Match             91.5%;  Score 1147.2;  DB 7;  Length 3400;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 1158;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3053 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 2994

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2993 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2934

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2933 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2874

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2873 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 2814

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2813 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2754

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2753 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 2694

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2693 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2634

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2633 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2574

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2573 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2514

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2513 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2454

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2453 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2394

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2393 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 2334

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2333 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2274

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2273 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2214

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2213 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2154

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2153 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 2094

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2093 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2034

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1974

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1914

Qy       1190 GAAAAGGATCCGTTTTCGCATTTATCGTGAAACGCT 1225
              |||||||||    |    ||   |  ||| |  ||
Db       1913 GAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT 1878


RESULT 3
ABN84078/c
ID   ABN84078 standard; DNA; 6122 BP.
XX
AC   ABN84078;
XX
DT   23-SEP-2002  (first entry)
XX
DE   Plasmid pRL5-CAT.
XX
KW   Plasmid pTL5-CAT; vector; antibody; gene; ds.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             591. .800
FT                   /*tag=  a
FT                   /partial
FT                   /product= "Amp frag"
FT                   /note= "the CDS does not include a stop codon"
FT   CDS             1013. .1672
FT                   /*tag=  b
FT                   /product= "Chloramphenicol transferase"
FT   misc_feature    complement(2052. .2651)
FT                   /*tag=  c
FT                   /note= "ori"
FT   promoter        2831. .3046
FT                   /*tag=  d
FT                   /note= "lac promoter"
FT   misc_signal     3009. .3029
FT                   /*tag=  e
FT                   /note= "lac rep site"
FT   RBS             3036. .3039
FT                   /*tag=  f
FT   CDS             3050. .3115
FT                   /*tag=  g
FT                   /partial
FT                   /product= "OmpA leader"
FT                   /note= "the CDS does not include a stop codon"
```

```
FT   misc_feature    3121. .3946
FT                   /*tag=  h
FT                   /note= "light chain variable region stuffer"
FT   misc_feature    3951. .4269
FT                   /*tag=  i
FT                   /note= "Kappa Cns"
FT   CDS             4298. .4363
FT                   /*tag=  j
FT                   /partial
FT                   /product= "pelB leader"
FT                   /note= "the CDS does not include a stop codon"
FT   misc_feature    4365. .5147
FT                   /*tag=  k
FT                   /note= "heavy chain variable region stuffer"
FT   CDS             5485. .5556
FT                   /*tag=  l
FT                   /partial
FT                   /product= "Linker-His tag-HA tag"
FT                   /note= "the CDS does not include a start codon"
FT   CDS             5557. .6090
FT                   /*tag=  m
FT                   /partial
FT                   /gene= "Gene III"
FT                   /note= "the CDS does not include a start codon"
XX
PN   WO200246435-A2.
XX
PD   13-JUN-2002.
XX
PF   05-DEC-2001; 2001WO-US047452.
XX
PR   05-DEC-2000; 2000US-0251440P.
XX
PA   (ALEX-) ALEXION PHARM INC.
XX
PI   Bowdish KS,  Barbas-Frederickson S,  Lin Y,  Renshaw M,  Wild M;
PI   Mcwhirter J;
XX
DR   WPI; 2002-537569/57.
XX
PT   Plasmid for in situ production of genes, comprises two template annealing
PT   sequences, such as downstream primer and upstream collar sequence and a
PT   restriction site located between the annealing sequences.
XX
PS   Disclosure; Fig 4A-T; 65pp; English.
XX
CC   The present sequence is that of plasmid pRL5-CAT, a derivative of pComb
CC   3X which has been modified to contain chloramphenicol resistance. The
CC   invention relates to the engineering of plasmids, of which pRL-CAT is a
CC   preferred example, for in situ production of genes. It was found that
CC   nucleic acids encoding a polypeptide can be directly incorporated into a
CC   plasmid by DNA polymerisation or by reverse transcription of a nucleic
CC   acid template. Preferably, nucleic acids encoding at least a portion of
CC   an antibody can be directly incorporated into the plasmid by reverse
CC   transcription of mRNA. The plasmids are engineered to contain 2 template
CC   annealing sequences (see ABN84079-80 and ABN84082-83), i.e. a downstream
CC   primer that anneals to a first portion of a nucleic acid template, e.g.
CC   mRNA encoding at least a portion of an antibody, an upstream collar
CC   sequence that anneals to a second portion of the template, and at least 1
CC   restriction site located between the 2 template annealing sequences. A
CC   single-stranded DNA plasmid vector is produced containing a nucleic acid
CC   encoding at least a portion of a polypeptide, e.g. a light chain and/or a
CC   heavy chain of an antibody. This vector can be transformed into a host
CC   cell and amplified
XX
SQ   Sequence 6122 BP; 1449 A; 1571 C; 1639 G; 1463 T; 0 U; 0 Other;

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6122;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802
```

```
Qy       1190 GAAAAG 1195
              ||||||
Db        801 GAAAAG 796


RESULT 4
AAL41112/c
ID   AAL41112 standard; DNA; 6122 BP.
XX
AC   AAL41112;
XX
DT   16-OCT-2002  (first entry)
XX
DE   Plasmid pRL5 CAT nucleic acid sequence.
XX
KW   Plasmid; vector; replication; expression; foreign genetic; bacteria;
KW   cyanobacterium; Escherichia coli; pComb3X plasmid; pRL5; pRL5 asc-CAT;
KW   pRL5-CAT; pRL5 bsi-CAT; gene; ds.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1013. .1672
FT                   /*tag= a
FT                   /product= "CAT amino acid sequence"
FT   CDS             3050. .3115
FT                   /*tag= b
FT                   /product= "Omp A leader amino acid sequence"
FT   CDS             3953. .4267
FT                   /*tag= c
FT                   /product= "Kappa constant region amino acid sequence"
FT   CDS             4298. .4363
FT                   /*tag= d
FT                   /product= "pel B leader amino acid sequence"
FT   CDS             5179. .6090
FT                   /*tag= e
FT                   /product= "CH1, His6 tag, HA tag, and gene III amino acid
FT                   sequence"
XX
PN   WO200246436-A2.
XX
PD   13-JUN-2002.
XX
PF   07-DEC-2001; 2001WO-US046516.
XX
PR   08-DEC-2000; 2000US-0254411P.
XX
PA   (ALEX-) ALEXION PHARM INC.
XX
PI   Bowdish KS,  Barbas-Frederickson S,  Wild M,  Mcwhirter J;
XX
DR   WPI; 2002-537570/57.
DR   P-PSDB; AAO22535, AAO22536, AAO22537, AAO22538, AAO22539.
XX
PT   Novel plasmid useful in cloning and expression of foreign genetic
PT   information.
XX
PS   Claim 1; Fig 7A-E; 39pp; English.
XX
CC   The invention relates to a plasmid with a nucleic acid sequence of 6122,
CC   6125, or 5683 base pairs fully defined in the specification. The
CC   invention more specifically relates to novel vectors capable of
CC   replication and expression of foreign genetic information in bacteria,
CC   such as, for example, cyanobacterium and E. coli. The new vectors have
CC   been designed to overcome certain drawbacks of the pComb3X plasmid. These
CC   new vectors include pRL5, pRL5-CAT, pRL5 asc-CAT and pRL5 bsi-CAT. This
CC   polynucleotide sequence represents the plasmid pRL5-CAT nucleic acid
CC   sequence of the invention
XX
SQ   Sequence 6122 BP; 1449 A; 1571 C; 1639 G; 1463 T; 0 U; 0 Other;

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6122;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy       1190 GAAAAG 1195
              ||||||
Db        801 GAAAAG 796


RESULT 5
AAL41115/c
ID   AAL41115 standard; DNA; 6122 BP.
XX
AC   AAL41115;
XX
DT   16-OCT-2002  (first entry)
XX
DE   Plasmid pRL5 bsi-CAT nucleic acid sequence.
XX
KW   Plasmid; vector; replication; expression; foreign genetic; bacteria;
KW   cyanobacterium; Escherichia coli; pComb3X plasmid; pRL5; pRL5 asc-CAT;
KW   pRL5-CAT; pRL5 bsi-CAT; gene; ds.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1013. .1672
FT                   /*tag=  a
FT                   /product= "CAT amino acid sequence"
FT   CDS             3050. .3115
FT                   /*tag=  b
FT                   /product= "Omp A leader amino acid sequence"
FT   CDS             3947. .4267
FT                   /*tag=  c
FT                   /product= "Kappa constant region amino acid sequence"
FT   CDS             4298. .4363
FT                   /*tag=  d
FT                   /product= "pel B leader amino acid sequence"
FT   CDS             5179. .6090
FT                   /*tag=  e
FT                   /product= "CH1, His6 tag, HA tag, and gene III amino acid
FT                   sequence"
XX
PN   WO200246436-A2.
XX
PD   13-JUN-2002.
XX
PF   07-DEC-2001; 2001WO-US046516.
XX
PR   08-DEC-2000; 2000US-0254411P.
XX
PA   (ALEX-) ALEXION PHARM INC.
XX
PI   Bowdish KS,  Barbas-Frederickson S,  Wild M,  Mcwhirter J;
XX
DR   WPI; 2002-537570/57.
DR   P-PSDB; AAO22535, AAO22536, AAO22538, AAO22539, AAO22541.
XX
PT   Novel plasmid useful in cloning and expression of foreign genetic
PT   information.
XX
PS   Disclosure; Fig 11A-E; 39pp; English.
XX
CC   The invention relates to a plasmid with a nucleic acid sequence of 6122,
CC   6125, or 5683 base pairs fully defined in the specification. The
CC   invention more specifically relates to novel vectors capable of
CC   replication and expression of foreign genetic information in bacteria,
CC   such as, for example, cyanobacterium and E. coli. The new vectors have
CC   been designed to overcome certain drawbacks of the pComb3X plasmid. These
CC   new vectors include pRL5, pRL5-CAT, pRL5 asc-CAT and pRL5 bsi-CAT. This
CC   polynucleotide sequence represents the plasmid pRL5 bsi-CAT nucleic acid
CC   sequence of the invention
XX
SQ   Sequence 6122 BP; 1447 A; 1571 C; 1641 G; 1463 T; 0 U; 0 Other;

  Query Match            91.1%;  Score 1142.8;  DB 6;  Length 6122;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy       1190 GAAAAG 1195
                ||||||
Db        801 GAAAAG 796


RESULT 6
AAL41114/c
ID   AAL41114 standard; DNA; 6125 BP.
XX
AC   AAL41114;
XX
DT   16-OCT-2002  (first entry)
XX
DE   Plasmid pRL5 CAT-Asc nucleic acid sequence.
XX
KW   Plasmid; vector; replication; expression; foreign genetic; bacteria;
KW   cyanobacterium; Escherichia coli; pComb3X plasmid; pRL5; pRL5 asc-CAT;
KW   pRL5-CAT; pRL5 bsi-CAT; gene; ds.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1013. .1672
FT                   /*tag=  a
FT                   /product= "CAT amino acid sequence"
FT   CDS             3050. .3115
FT                   /*tag=  b
FT                   /product= "Omp A leader amino acid sequence"
FT   CDS             3953. .4270
FT                   /*tag=  c
FT                   /product= "Kappa constant region amino acid sequence"
FT   CDS             4301. .4366
FT                   /*tag=  d
FT                   /product= "pel B leader amino acid sequence"
FT   CDS             5182. .6093
FT                   /*tag=  e
FT                   /product= "CH1, His6 tag, HA tag, and gene III amino acid
FT                   sequence"
XX
PN   WO200246436-A2.
XX
PD   13-JUN-2002.
XX
PF   07-DEC-2001; 2001WO-US046516.
XX
PR   08-DEC-2000; 2000US-0254411P.
XX
PA   (ALEX-) ALEXION PHARM INC.
XX
PI   Bowdish KS,  Barbas-Frederickson S,  Wild M,  Mcwhirter J;
XX
DR   WPI; 2002-537570/57.
DR   P-PSDB; AAO22535, AAO22536, AAO22538, AAO22539, AAO22540.
XX
PT   Novel plasmid useful in cloning and expression of foreign genetic
PT   information.
XX
PS   Claim 1; Fig 9A-E; 39pp; English.
XX
CC   The invention relates to a plasmid with a nucleic acid sequence of 6122,
CC   6125, or 5683 base pairs fully defined in the specification. The
CC   invention more specifically relates to novel vectors capable of
CC   replication and expression of foreign genetic information in bacteria,
CC   such as, for example, cyanobacterium and E. coli. The new vectors have
CC   been designed to overcome certain drawbacks of the pComb3X plasmid. These
CC   new vectors include pRL5, pRL5-CAT, pRL5 asc-CAT and pRL5 bsi-CAT. This
CC   polynucleotide sequence represents the plasmid pRL5 CAT-Asc nucleic acid
CC   sequence of the invention
XX
SQ   Sequence 6125 BP; 1447 A; 1573 C; 1644 G; 1461 T; 0 U; 0 Other;

  Query Match             91.1%;  Score 1142.8;  DB 6;  Length 6125;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1144;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
                |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1941 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 1882

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1881 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 1822

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1821 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 1762

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1761 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 1702

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1701 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 1642

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1641 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 1582

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1581 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 1522

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1521 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 1462

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1461 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 1402

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1401 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 1342

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1341 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 1282

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1281 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 1222

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1221 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 1162

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1161 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1102

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1101 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1042
```

```
Qy       950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1041 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 982

Qy      1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       981 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 922

Qy      1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       921 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 862

Qy      1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       861 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 802

Qy      1190 GAAAAG 1195
            ||||||
Db       801 GAAAAG 796
```

RESULT 7
AEA43207/c
ID AEA43207 standard; DNA; 4100 BP.
XX
AC AEA43207;
XX
DT 25-AUG-2005 (first entry)
XX
DE M. smegmatis perhydrolase expression vector pAH502R.
XX
KW perhydrolase; bleaching; paper; pulp; cosmetics; vector; ds.
XX
OS Mycobacterium smegmatis.
OS Synthetic.
XX
PN WO2005056782-A2.
XX
PD 23-JUN-2005.
XX
PF 03-DEC-2004; 2004WO-US040438.
XX
PR 03-DEC-2003; 2003US-0526764P.
XX
PA (GEMV) GENENCOR INT INC.
PA (PROC) PROCTER & GAMBLE CO.
XX
PI Amin NS, Boston MG, Bott RR, Cervin MA, Concar EM, Gustwiller ME;
PI Jones BE, Liebeton K, Miracle GS, Oh H, Poulose AJ, Ramer SW;
PI Scheibel JJ, Weyler W, Whited GM;
XX
DR WPI; 2005-458588/46.
XX
PT New isolated perhydrolase having preset perhydrolysis to hydrolysis
PT ratio, useful in cleaning, bleaching and disinfecting applications e.g.
PT paper pulp bleaching.
XX
PS Example 6; SEQ ID NO 133; 523pp; English.
XX
CC The invention relates to an isolated perhydrolase (I), exhibiting a
CC perhydrolysis to hydrolysis ratio that is greater than 1. A composition
CC containing (I) is useful for cleaning. (I) is useful in cleaning,
CC bleaching and disinfecting applications such as paper and pulp bleaching,
CC fabric and garment cleaning, hard surface cleaning and personal care
CC applications (e.g. oral care, hair care, and skin care). (I) is useful in
CC applications where peracid hydrolysis is useful, such as in the clean up
CC of peracids. (I) enables effective cleaning, bleaching and disinfecting
CC over broad pH and temperature ranges. The present sequence represents a
CC M. smegmatis perhydrolase expression vector.
XX
SQ Sequence 4100 BP; 996 A; 1044 C; 1090 G; 970 T; 0 U; 0 Other;

Query Match 90.5%; Score 1135.2; DB 14; Length 4100;

Best Local Similarity 98.4%; Pred. No. 0;
Matches 1157; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

```
Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3753 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 3694

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3693 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 3634

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3633 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 3574

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3573 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 3514

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3513 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 3454

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3453 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 3394

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3393 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 3334

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3333 AC-GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 3275

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3274 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 3215

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3214 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 3155

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3154 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 3095

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3094 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 3035

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3034 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2975

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2974 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2915

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2914 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2855

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2854 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 2795

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2794 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 2735

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       2734 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 2675

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2674 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 2615

Qy       1190 GAAAAGGATCCGTTTTCGCATTTATCGTGAAACGCT 1225
              |||||||||    |    ||   |  ||| |  ||
Db       2614 GAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCT 2579


RESULT 8
ADD14893/c
ID   ADD14893 standard; DNA; 3817 BP.
XX
AC   ADD14893;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Phage display vector pABMbd-1.
XX
KW   ds; adapter-directed display system; phage display;
KW   outer surface protein; homodimerisation domain;
KW   heterodimerisation domain; antigen-binding unit; cell surface receptor;
KW   receptor ligand; cytosolic protein; secreted protein; HA tag; OmpA;
KW   GABAbR; GR2; Lpp.
XX
OS   Synthetic.
OS   Escherichia coli.
OS   Enterobacteria phage M13.
OS   Homo sapiens.
XX
PN   US2003104355-A1.
XX
PD   05-JUN-2003.
XX
PF   02-NOV-2001; 2001US-00033399.
XX
PR   02-NOV-2001; 2001US-00033399.
XX
PA   (WANG/) WANG C.
PA   (ZHON/) ZHONG P.
PA   (WANG/) WANG X.
XX
PI   Wang C,  Zhong P,  Wang X;
XX
DR   WPI; 2003-801222/75.
XX
PT   Adapter-directed display system having expression vector with sequence
PT   encoding exogenous polypeptide and helper vector, for displaying
PT   exogenous polypeptide e.g., receptor ligand on outer surface of genetic
PT   package.
XX
PS   Example 6; SEQ ID NO 24; 77pp; English.
XX
CC   The invention relates to an adapter-directed display system for
CC   displaying an exogenous polypeptide on the outer surface of a genetic
CC   package, comprising an expression vector with a sequence that encodes
CC   exogenous polypeptide fused to a first adapter sequence, and a helper
CC   vector with outer-surface sequences encoding proteins fused to a second
CC   adapter, the polypeptide is produced in a host cell to cause display of
CC   the polypeptide. Also included are a helper vector for displaying a
CC   polypeptide on the outer surface of a genetic package (comprising, outer-
CC   surface sequences necessary for packaging the genetic package, where at
CC   least one of the surface presenting sequences is fused in-frame to an
CC   adapter, the adapter acting, when the polypeptide is produced in a
CC   suitable host cell, to cause the display of the polypeptide), an
CC   expression vector for producing a polypeptide within or on the outer
CC   surface of a genetic package (comprising a coding sequence encoding the
CC   polypeptide fused in-frame to a first adapter, where the vector is devoid
CC   of outer-surface sequences encoding any functional outer-surface proteins
CC   of the genetic package, and expression of the polypeptide on the outer
CC   surface of the genetic package is mediated through non-covalent pairwise
CC   interaction between the first adapter and a second adapter, where the
CC   second adapter is fused to an outer-surface protein), a kit comprising
CC   above vectors in suitable packaging, a host cell comprising the vectors,
CC   a polypeptide displayed on the outer surface of a genetic package using
CC   the adapter-directed display system, a genetic package displaying on its
CC   outer surface the fusion polypeptide, a selectable library comprising
CC   several genetic packages, at least one being the expression vector, a
CC   selectable library comprising several genetic packages (at least one
CC   member of several packages displaying a polypeptide on its outer surface
CC   using the adapter-directed display system) and detecting the presence of
CC   a specific interaction between a test agent and an exogenous polypeptide
CC   that is displayed on a genetic package (involving providing a genetic
CC   package displaying the exogenous polypeptide that is prepared using the
CC   adapter-directed display system, contacting the genetic package with the
CC   test agent and detecting the formation of the stable polypeptide-agent
CC   complex on the genetic package). The outer-surface sequences are chosen
CC   from gene III, gene VI, gene VII, gene VIII and gene IX of a filamentous
CC   phage. The outer-surface sequences encode bacterial outer-surface
CC   proteins chosen from Lpp-OmpA, TraT, Pal, Oprl, Inp and AIDA-I. The first
CC   and second adapters are homodimerisation or heterodimerisation sequences.
CC   The adapter-directed display system is useful for displaying a
CC   polypeptide on the outer surface of a genetic package, which involves
CC   causing the adapter-directed display system to be transcribed and
CC   translated in a suitable host cell. The selectable library is useful for
CC   obtaining a polypeptide with desired property, which involves providing
CC   the selectable library and screening the selectable library to obtain at
CC   least one genetic package displaying a polypeptide with the desired
CC   property. The desired property is binding specificity to an agent of
CC   interest. Screening the selectable library further involves isolating the
CC   genetic package that displays a polypeptide having the desired property.
CC   Isolating the genetic package further involves obtaining a nucleotide
CC   sequence from the genetic package that encodes the polypeptide with the
CC   desired property. The polypeptide with the desired property is chosen
CC   from antigen-binding unit, cell surface receptor, receptor ligand,
CC   cytosolic protein, secreted protein, nuclear protein, and their
CC   functional motif. The present sequence represents a vector of the
CC   invention which can express an exogenous protein as a fusion protein with
CC   an Lpp-OmpA-GR2 fusion protein using GABAb receptor GR2 domain as the
CC   adapter sequence.
XX
SQ   Sequence 3817 BP; 948 A; 930 C; 948 G; 991 T; 0 U; 0 Other;

  Query Match             90.2%;  Score 1130.8;  DB 10;  Length 3817;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 1143;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              |||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       2834 AAAAAGGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 2775

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2774 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2715

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2714 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2655

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2654 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 2595

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2594 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2535

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2534 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 2475

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2474 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2415
```

```
Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2414 AC-GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2356

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2355 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2296

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2295 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2236

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2235 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2176

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2175 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 2116

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2115 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2056

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2055 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 1996

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1995 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 1936

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1935 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1876

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1875 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1816

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1815 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1756

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1755 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1696

Qy       1190 GAAAAG 1195
              ||||||
Db       1695 GAAAAG 1690
```

RESULT 9
ADZ03868/c
ID ADZ03868 standard; DNA; 6607 BP.
XX
AC ADZ03868;
XX
DT 16-JUN-2005 (first entry)
XX
DE Plasmid P15a DNA sequence #1.
XX
KW Pharmaceutical; P15a; ds.
XX
OS Plasmid P15a.
XX
PN WO2005030972-A1.
XX
PD 07-APR-2005.
XX
PF 29-SEP-2004; 2004WO-JP014213.
XX
PR 29-SEP-2003; 2003JP-00337663.
XX
PA (JAPC) NIPPON SHOKUBAI CO LTD.
XX
PI Toraya T, Tobimatsu T, Yamanishi M, Mori K, Kajiura H, Yamada S;
PI Yuzuki M, Azuma M, Hara T, Yasuda S;
XX
DR WPI; 2005-315044/32.
XX
PT Manufacture of 3-hydroxy propionaldehyde useful as raw material for
PT manufacturing 1,3-propanediol, by dehydrating glycerol with diol
PT dehydratase and/or glycerol dehydratase and microbial cells.
XX
PS Example 1; SEQ ID NO 2; 40pp; Japanese.
XX
CC The invention relates to a method for manufacturing 3-
CC hydroxypropionaldehyde involving dehydrating glycerol with diol
CC dehydratase and/or glycerol dehydratase and microbial cells comprising
CC diol dehydratase reactivation factor and/or glycerol dehydratase
CC reactivation factor. The invention also relates to a method for
CC manufacturing 1,3-propanediol, a method for manufacturing 3-
CC hydroxypropionic acid and methods for manufacturing acrolein, acrylic
CC acid and acrylic ester. The dehydration reaction of glycerol is performed
CC under aerobic conditions using microbial cells. This sequence represents
CC plasmid P15a DNA used in the method of the invention.
XX
SQ Sequence 6607 BP; 1518 A; 1801 C; 1831 G; 1457 T; 0 U; 0 Other;

Query Match 89.0%; Score 1116; DB 14; Length 6607;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1136; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

```
Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 4802

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 4742

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 4682

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4681 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGC-ACCAGGCGTT 4623

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4622 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 4563

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       4562 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 4503

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4502 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 4443

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4442 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 4383

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4382 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 4323

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4322 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 4263
```

```
Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4262 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 4203

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       4202 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGA 4143

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4142 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 4083

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4082 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 4023

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4022 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 3963

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3962 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 3903

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3902 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 3843

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3842 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 3783

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3782 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 3723

Qy       1190 GAAAAGGATCCG 1201
                 |||  |  ||
Db       3722 CAAAGTGCGTCG 3711
```

RESULT 10
ADZ03870/c
ID ADZ03870 standard; DNA; 6608 BP.
XX
AC ADZ03870;
XX
DT 16-JUN-2005 (first entry)
XX
DE Plasmid P15a DNA sequence #2.
XX
KW Pharmaceutical; P15a; ds.
XX
OS Plasmid P15a.
XX
PN WO2005030972-A1.
XX
PD 07-APR-2005.
XX
PF 29-SEP-2004; 2004WO-JP014213.
XX
PR 29-SEP-2003; 2003JP-00337663.
XX
PA (JAPC) NIPPON SHOKUBAI CO LTD.
XX
PI Toraya T, Tobimatsu T, Yamanishi M, Mori K, Kajiura H, Yamada S;
PI Yuzuki M, Azuma M, Hara T, Yasuda S;
XX
DR WPI; 2005-315044/32.
XX
PT Manufacture of 3-hydroxy propionaldehyde useful as raw material for
PT manufacturing 1,3-propanediol, by dehydrating glycerol with diol
PT dehydratase and/or glycerol dehydratase and microbial cells.
XX
PS Example 2; SEQ ID NO 4; 40pp; Japanese.
XX
CC The invention relates to a method for manufacturing 3-
CC hydroxypropionaldehyde involving dehydrating glycerol with diol
CC dehydratase and/or glycerol dehydratase and microbial cells comprising
CC diol dehydratase reactivation factor and/or glycerol dehydratase
CC reactivation factor. The invention also relates to a method for
CC manufacturing 1,3-propanediol, a method for manufacturing 3-
CC hydroxypropionic acid and methods for manufacturing acrolein, acrylic
CC acid and acrylic ester. The dehydration reaction of glycerol is performed
CC under aerobic conditions using microbial cells. This sequence represents
CC plasmid P15a DNA used in the method of the invention.
XX
SQ Sequence 6608 BP; 1495 A; 1749 C; 1918 G; 1446 T; 0 U; 0 Other;

Query Match 89.0%; Score 1116; DB 14; Length 6608;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1136; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

```
Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4862 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 4803

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4802 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 4743

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4742 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 4683

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       4682 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGC-ACCAGGCGTT 4624

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4623 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 4564

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       4563 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 4504

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4503 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 4444

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4443 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 4384

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4383 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 4324

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4323 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 4264

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4263 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 4204

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       4203 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGA 4144

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4143 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 4084

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4083 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 4024

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4023 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 3964

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3963 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 3904

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3903 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 3844

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3843 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 3784

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3783 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 3724

Qy       1190 GAAAAGGATCCG 1201
                |||  |   ||
Db       3723 CAAAGTGCGTCG 3712
```

RESULT 11
ADZ66224/c
ID ADZ66224 standard; cDNA; 9808 BP.
XX
AC ADZ66224;
XX
DT 14-JUL-2005 (first entry)
XX
DE Construct pAM401-MCS.
XX
KW ss; Cytostatic; Antiasthmatic; Respiratory-Gen; Antipsoriatic;
KW Dermatological; Antiangiogenic; Ophthalmological; Antidiabetic;
KW Antirheumatic; Antiarthritic; Antiinflammatory; Immunosuppressive;
KW Uropathic; Gynecological; Antiatherosclerotic; Cardiant; Vasotropic;
KW Vaccine; EphA2; receptor tyrosine kinase; Ephrin; hyperproliferation;
KW Listeria; angiogenesis; cancer; skin; lung; colon; breast; prostate;
KW bladder; pancreas; renal cell carcinoma; melanoma; leukemia; lymphoma;
KW asthma; chronic pulmonary obstructive disease; lung fibrosis;
KW bronchial hyper responsiveness; psoriasis; seborrhoeic dermatitis;
KW macular degeneration; diabetic retinopathy; retinopathy of prematurity;
KW vascular restenosis; infantile hemangioma; verruca vulgaris; psoriasis;
KW Kaposi's sarcoma; neurofibromatosis;
KW recessive dystrophic epidermolysis bullosa; rheumatoid arthritis;
KW ankylosing spondylitis; systemic lupus; psoriatic arthropathy;
KW Reiters' syndrome; Sjogrens' syndrome; endometriosis; preeclampsia;
KW atherosclerosis; coronary artery disease; gene.
XX
OS Synthetic.
XX
PN WO2005037233-A2.
XX
PD 28-APR-2005.
XX
PF 15-OCT-2004; 2004WO-US034694.
XX
PR 15-OCT-2003; 2003US-0511719P.
PR 15-OCT-2003; 2003US-0511919P.
PR 24-DEC-2003; 2003US-0532666P.
PR 26-MAR-2004; 2004US-0556631P.
PR 01-OCT-2004; 2004US-0615470P.
PR 07-OCT-2004; 2004US-0617544P.
XX
PA (MEDI-) MEDIMMUNE INC.
PA (CERU-) CERUS CORP.
XX
PI Kinch MS, Kiener PA, Bruckheimer E, Dubensky TW, Cook DN;
XX
DR WPI; 2005-322763/33.
XX
PT Eliciting an immune response against EphA2-expressing cells in subjects,
PT by administering composition comprising Listeria bacterium that expresses
PT EphA2 antigenic peptide, to subject.
XX
PS Disclosure; SEQ ID NO 43; 219pp; English.
XX
CC This sequence represents the expression cassette construct pAM401-MCS.
CC This sequence was used in the scope of the invention for the expression
CC of EphA2. EphA2 is a 130 kD receptor tyrosine kinase that is expressed at
CC low levels in adult epithelia and is enriched within sites of cell-cell
CC adhesion. It binds ligands (Ephrins A1-A5) that are anchored to the cell
CC membrane. Ligand binding causes autophosphorylation, however EphA2
CC retains enzymatic activity in the absence of ligand binding or
CC phosphotyrosine content. EphA2 is upregulated on a large number of
CC hyperproliferation cells, including aggressive carcinoma cells. The
CC method of the invention for eliciting an immune response against an EphA2
CC -expressing cell in a subject, involves administering to a subject a
CC composition comprising a Listeria bacterium that expresses an EphA2
CC antigenic peptide, to elicit an immune response against an EphA2-
CC expressing cell. The invention contains further methods for treating a
CC human subject having a hyperproliferative disorder of EphA2-expressing
CC cells, involving administering to the subject a composition comprising
CC EphA2 antigen peptide-expressing Listeria bacterium, to treat a
CC hyperproliferative disorder of EphA2-expressing cells; and treating a
CC human subject having a disease involving aberrant angiogenesis involving
CC administering to the subject a composition comprising EphA2 antigenic
CC peptide-expressing Listeria bacterium, to treat disease involving
CC aberrant angiogenesis. The method of the invention is useful for
CC eliciting an immune response against an EphA2-expressing cell in a
CC subject. The subject has cancer. The cancer is of an epithelial cell
CC origin or T cell origin. The cancer is a cancer of skin, lung, colon,
CC breast, prostate, bladder or pancreas or is a renal cell carcinoma or
CC melanoma. The cancer is a leukemia or a lymphoma. The subject has a non-
CC neoplastic hyperproliferative disorder. The hyperproliferative disorder
CC is an epithelial cell disorder. The hyperproliferative disorder is
CC asthma, chronic pulmonary obstructive disease, lung fibrosis, bronchial
CC hyper responsiveness, psoriasis, and seborrhoeic dermatitis. The second
CC method is useful for treating a human subject having a hyperproliferative
CC disorder of EphA2-expressing cells. The third method is useful for
CC treating a subject having a disease involving aberrant angiogenesis. The
CC disease is macular degeneration, diabetic retinopathy, retinopathy of
CC prematurity, vascular restenosis, infantile hemangioma, verruca vulgaris,
CC psoriasis, Kaposi's sarcoma, neurofibromatosis, recessive dystrophic
CC epidermolysis bullosa, rheumatoid arthritis, ankylosing spondylitis,
CC systemic lupus, psoriatic arthropathy, Reiters' syndrome, and Sjogrens'
CC syndrome, endometriosis, preeclampsia, atherosclerosis or coronary artery
CC disease.
XX
SQ Sequence 9808 BP; 2836 A; 1923 C; 1946 G; 3103 T; 0 U; 0 Other;

```
  Query Match             89.0%;  Score 1116;  DB 14;  Length 9808;
  Best Local Similarity   98.6%;  Pred. No. 0;
  Matches 1136;  Conservative    0;  Mismatches   15;  Indels    1;  Gaps    1;

Qy          50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
               | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9192 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 9133

Qy         110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9132 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 9073

Qy         170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        9072 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 9013

Qy         230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        9012 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGC-ACCAGGCGTT 8954
```

```
Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8953 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 8894

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       8893 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 8834

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8833 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 8774

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8773 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 8714

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8713 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 8654

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8653 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 8594

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8593 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 8534

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       8533 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 8474

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8473 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 8414

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8413 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 8354

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8353 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 8294

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8293 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 8234

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8233 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 8174

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8173 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 8114

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8113 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 8054

Qy       1190 GAAAAGGATCCG 1201
              ||| |  | ||
Db       8053 CAAAGTGCGTCG 8042


RESULT 12
AEB80052/c
ID   AEB80052 standard; DNA; 9808 BP.
XX
AC   AEB80052;
XX
DT   06-OCT-2005  (first entry)
XX
DE   Vector pAM401/pPL2 containing hly promoter/p60 sig pep. cassette.
XX
KW   Vector; ds; listeriolysin O; Cytostatic; Antimicrobial;
KW   Immunosuppressive; Immunostimulant; vaccine; fusion protein;
KW   protein production; protein secretion; immune response; cancer;
KW   infection; autoimmune disease.
XX
OS   Listeria monocytogenes.
OS   Synthetic.
OS   Unidentified.
XX
PN   WO2005071088-A2.
XX
PD   04-AUG-2005.
XX
PF   23-DEC-2004; 2004WO-US044080.
XX
PR   24-DEC-2003; 2003US-0532598P.
PR   02-FEB-2004; 2004US-0541515P.
PR   06-FEB-2004; 2004US-00773618.
PR   06-FEB-2004; 2004US-00773792.
PR   26-MAR-2004; 2004US-0556744P.
PR   30-JUN-2004; 2004US-00883599.
PR   23-JUL-2004; 2004WO-US023881.
PR   05-AUG-2004; 2004US-0599377P.
PR   01-OCT-2004; 2004US-0615287P.
PR   06-OCT-2004; 2004US-0616750P.
XX
PA   (CERU-) CERUS CORP.
XX
PI   Dubensky TW,  Portnoy DA,  Luckett WS,  Cook DN;
XX
DR   WPI; 2005-534010/54.
XX
PT   New recombinant nucleic acid molecule (comprising a first polynucleotide
PT   encoding a signal peptide native to a bacterium; and a second
PT   polynucleotide encoding a polypeptide) useful e.g. to induce immune
PT   response in host to antigen.
XX
PS   Example 13; SEQ ID NO 33; 271pp; English.
XX
CC   The invention relates to a recombinant nucleic acid molecule (I)
CC   (comprising a first polynucleotide (A) encoding a signal peptide native
CC   to a bacterium, where (A) is codon-optimized for expression in the
CC   bacterium, and a second polynucleotide (B) encoding a polypeptide, where
CC   (B) is in the same translational reading frame as (A), where (A) encodes
CC   a fusion protein comprising the signal peptide and the polypeptide). Also
CC   included are the expression cassette (I) further comprising a promoter
CC   operably linked to (A) and (B), a recombinant Listeria bacterium (II)
CC   comprising (I), an immunogenic composition or vaccine comprising (II), a
CC   method of preventing or treating a condition in a host (comprising
CC   administering the recombinant Listeria bacterium to the host), a
CC   recombinant Listeria bacterium comprising a polycistronic expression
CC   cassette. The polypeptide encoded by (B) comprises an antigen (a tumor-
CC   associated antigen, a polypeptide derived from a tumor-associated
CC   antigen, an infectious disease antigen, or a polypeptide derived from an
CC   infectious disease antigen). The polypeptide encoded by (B) is
CC   heterologous to the signal peptide, foreign to the Listeria bacterium, or
CC   both. The signal peptide is an LLO (listeriolysin O, encoded by the hly
CC   gene) signal peptide from Listeria monocytogenes or is a p60 signal
CC   peptide from Listeria monocytogenes. The heterologous protein is
CC   preferably mesothelin, an antigenic fragment or its variants, NY-ESO-1,
CC   or an antigenic fragment or its variant). The polypeptide encoded by (B)
CC   comprises human mesothelin deleted of its signal peptide and GPI linker
CC   domain. The recombinant bacterium is useful in the manufacture of a
CC   medicament for inducing an immune response in a host to an antigen, and
CC   for preventing or treating a condition in a host (where the condition is
CC   cancer, infections and autoimmunity. The recombinant nucleic acid
CC   molecule, expression cassettes and vectors are useful for expressing
CC   and/or secreting polypeptides including heterologous polypeptides such as
CC   antigens in bacteria. The present sequence represents the Vector pAM401
CC   which contains the hly promoter/p60 signal peptide cassette.
```

```
XX
SQ   Sequence 9808 BP; 2836 A; 1923 C; 1946 G; 3103 T; 0 U; 0 Other;

  Query Match            89.0%;  Score 1116;  DB 14;  Length 9808;
  Best Local Similarity  98.6%;  Pred. No. 0;
  Matches 1136;  Conservative    0;  Mismatches   15;  Indels    1;  Gaps    1;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
              | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9192 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 9133

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9132 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 9073

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9072 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 9013

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       9012 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGC-ACCAGGCGTT 8954

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8953 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 8894

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       8893 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 8834

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8833 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 8774

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8773 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 8714

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8713 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 8654

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8653 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 8594

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8593 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 8534

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       8533 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 8474

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8473 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 8414

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8413 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 8354

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8353 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 8294

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8293 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 8234

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8233 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 8174

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8173 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 8114

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8113 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 8054

Qy       1190 GAAAAGGATCCG 1201
                 |||  |  ||
Db       8053 CAAAGTGCGTCG 8042


RESULT 13
AEB72933/c
ID   AEB72933 standard; DNA; 9808 BP.
XX
AC   AEB72933;
XX
DT   06-OCT-2005  (first entry)
XX
DE   DNA sequence for plasmid pAM401-MCS.
XX
KW   immune response; receptor tyrosine kinase; EphA2; antigen; expression;
KW   antibody production; adoptive immunotherapy; vaccine; hyperproliferation;
KW   cancer; angiogenesis disorder; immunostimulant; cytostatic;
KW   antiangiogenic; antiasthmatic; respiratory-gen; antipsoriatic;
KW   antiseborrheic; dermatological; ophthalmological; antidiabetic;
KW   antirheumatic; antiarthritic; antiinflammatory; gynecological;
KW   angiogenesis-inhibitor; vasotropic; neuroprotective; immunosuppressive;
KW   antiarteriosclerotic; cardiant; pAM401-MCS; ds.
XX
OS   Synthetic.
OS   Unidentified.
XX
PN   WO2005067460-A2.
XX
PD   28-JUL-2005.
XX
PF   15-OCT-2004; 2004WO-US034693.
XX
PR   24-DEC-2003; 2003US-0532696P.
PR   18-AUG-2004; 2004US-0602588P.
PR   01-OCT-2004; 2004US-0615548P.
PR   07-OCT-2004; 2004US-0617564P.
XX
PA   (MEDI-) MEDIMMUNE INC.
XX
PI   Kinch MS;
XX
DR   WPI; 2005-522701/53.
XX
PT   Eliciting an immune response comprises administering to an individual a
PT   composition comprising an EphA2 antigenic peptide, an EphA2 antigenic
PT   peptide expression vehicle, or antigen presenting cells.
XX
PS   Disclosure; SEQ ID NO 43; 241pp; English.
XX
CC   The invention relates to a method of eliciting an immune response against
CC   an EphA2-expressing cell. The method comprises administering to an
CC   individual a composition comprising an EphA2 antigenic peptide, an EphA2
CC   antigenic peptide expression vehicle, antigen presenting cells sensitized
CC   with an EphA2 antigenic peptide, or an anti-idiotypic antibody or its
CC   antigen-binding fragment, which immunospecifically binds to an idiotype
CC   of an anti-EphA2 antibody, in an amount to elicit an immune response
CC   against an EphA2-expressing cell. Also described are: (1) a method of
CC   treating a human individual having a hyperproliferative disorder of an
CC   EphA2-expressing cell, or a disease involving aberrant angiogenesis,
CC   comprising performing the method of the invention; and (2) a method of
CC   producing antibodies that immunospecifically bind to EphA2, and
CC   administering them to an individual as an EphA2 vaccine. In the method of
```

CC the invention, the EphA2 antigenic peptide is NOT any of the sequences
CC given as SEQ ID Nos 3-18 in the specification. The composition further
CC comprises an adjuvant, a heat shock protein bound to the EphA2 antigenic
CC peptide, and expression vehicles expressing the EphA2 antigenic peptides.
CC The expression vehicle is an infectious agent comprising a nucleic acid,
CC the nucleic acid comprising a nucleotide sequence encoding the EphA2
CC antigenic peptide operably linked to a promoter, where the sequence
CC encoding the EphA2 antigenic peptide is codon-optimized for expression in
CC the infectious agent. The method of the invention is useful for treating
CC a human individual having a hyperproliferative disorder of EphA2-
CC expressing cells by administering to the individual a composition
CC comprising antibodies produced by administering an EphA2 vaccine to a
CC host in an amount to treat a hyperproliferative disorder of EphA2-
CC expressing cells. Alternatively, the individual can be administered with
CC a composition comprising an EphA2 expression vehicle, e.g. bacterium or
CC virus, in an amount to treat a hyperproliferative disorder of EphA2-
CC expressing cells, and administering to the individual an antibiotic or
CC antiviral agent in an amount to treat a bacterial or viral infection. The
CC methods and compositions of the invention are useful for eliciting an
CC immune response against an EphA2-expressing cell, and for treating a
CC hyperproliferative disorder such as cancer (e.g. cancer of the skin,
CC lung, colon, ovary, esophagus, breast, prostate, bladder or pancreas or
CC renal cell carcinoma or melanoma), a non-neoplastic hyperproliferative
CC disorder (e.g. an epithelial cell disorder, specifically asthma, chronic
CC pulmonary obstructive disease, lung fibrosis, bronchial hyper
CC responsiveness, psoriasis, or seborrheic dermatitis), or a disease
CC involving aberrant angiogenesis (e.g. macular degeneration, diabetic
CC retinopathy, retinopathy of prematurity, vascular restenosis, infantile
CC hemangioma, verruca vulgaris, Kaposi's sarcoma, neurofibromatosis,
CC recessive dystrophic epidermolysis bullosa, rheumatoid arthritis,
CC ankylosing spondylitis, systemic lupus, psoriatic arthropathy, Reiter's
CC syndrome, and Sjogren's syndrome, endometriosis, preeclampsia,
CC atherosclerosis or coronary artery disease). This sequence represents a
CC plasmid containing a multiple cloning site (MCS). It is used in the
CC exemplification of the method of the invention.
XX
SQ Sequence 9808 BP; 2836 A; 1923 C; 1946 G; 3103 T; 0 U; 0 Other;

Query Match 89.0%; Score 1116; DB 14; Length 9808;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1136; Conservative 0; Mismatches 15; Indels 1; Gaps 1;

```
Qy        50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
             | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9192 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 9133

Qy       110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9132 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 9073

Qy       170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9072 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 9013

Qy       230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db      9012 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGC-ACCAGGCGTT 8954

Qy       290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8953 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 8894

Qy       350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      8893 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 8834

Qy       410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8833 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 8774

Qy       470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8773 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 8714

Qy       530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8713 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 8654

Qy       590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8653 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 8594

Qy       650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8593 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 8534

Qy       710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db      8533 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 8474

Qy       770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8473 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 8414

Qy       830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8413 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 8354

Qy       890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8353 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 8294

Qy       950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8293 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 8234

Qy      1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8233 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 8174

Qy      1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8173 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 8114

Qy      1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8113 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 8054

Qy      1190 GAAAAGGATCCG 1201
             ||| |  ||
Db      8053 CAAAGTGCGTCG 8042
```

RESULT 14
AAV28398
ID AAV28398 standard; DNA; 5838 BP.
XX
AC AAV28398;
XX
DT 24-JUL-1998 (first entry)
XX
DE Plasmid pRZTL1 nucleotide sequence.
XX
KW Tn5 transposase; modified; enzyme; in vitro transposition; mutant;
KW target; marker; transposon 5; plasmid pRZTL1; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT insertion_seq 1. .19
FT /*tag= a
FT CDS 77. .1267
FT /*tag= b
FT /note= "tetracycline resistance"
FT CDS complement(2301. .2960)
FT /*tag= c

```
FT                   /note= "chloramphenicol resistance"
FT   insertion_seq   4564. .4582
FT                   /*tag=  d
FT   CDS             4715. .5530
FT                   /*tag=  e
FT                   /note= "kanamycin resistance"
XX
PN   WO9810077-A1.
XX
PD   12-MAR-1998.
XX
PF   09-SEP-1997;   97WO-US015941.
XX
PR   09-SEP-1996;   96US-00814877.
PR   02-MAY-1997;   97US-00850880.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Reznikoff WS,  Goryshin IY,  Zhou H;
XX
DR   WPI; 1998-193627/17.
DR   P-PSDB; AAW56695, AAW56696, AAW56697.
XX
PT   Modified Tn5 transposase construct used in novel system for in vitro
PT   transposition - used to, e.g. create absolute defective mutants, provide
PT   selective markers and to facilitate insertion of specialised DNA
PT   sequences into target DNA.
XX
PS   Disclosure; Page 35-46; 73pp; English.
XX
CC   This is the nucleotide sequence of the plasmid pRZTL1 which is used to
CC   demonstrate in vitro transposition of a transposable element located
CC   between a pair of Tn5 (transposon 5) outside end (OE) termini. The
CC   invention provides a genetic construct that contains a nucleotide
CC   sequence encoding a modified Tn5 transposase enzyme that has both greater
CC   avidity for Tn5 OE repeats and is less likely to assume an inactive
CC   multimeric form than a wild type Tn5 transposase and a transposable DNA
CC   sequence flanked at its 5' and 3' ends by an 18 or 19 base pair flanking
CC   DNA sequence comprising nucleotide A at position 10, T at 11 and A at 12.
CC   The modified Tn5 transposase and the transposable DNA which is a DNA
CC   donor molecule are used in a system for in vitro transposition. The
CC   system and method can be used to create absolute defective mutants, to
CC   provide selective markers to target DNA, to provide portable regions of
CC   homology to a target DNA, to facilitate insertion of specialised DNA
CC   sequences into target DNA, to provide primer binding sites or tags for
CC   DNA sequencing , to facilitate production of genetic fusion for gene
CC   expression studies and protein domain mapping, as well as to bring
CC   together other desired combinations of DNA sequences (combinatorial
CC   genetics). The modified Tn5 transposase facilitates in vitro
CC   transposition reaction rates of at least about 100-fold higher than can
CC   be achieved using wild type transposase (as measure in vivo). In vitro
CC   transposition using this system can also use donor DNA and target DNA
CC   that is circular or linear. The system also requires no outside high
CC   energy source and no other protein other than the modified transposase
XX
SQ   Sequence 5838 BP; 1370 A; 1525 C; 1516 G; 1427 T; 0 U; 0 Other;

  Query Match             88.9%;  Score 1114.4;  DB 2;  Length 5838;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 1135;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy          50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
               | ||  |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2033 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 2092

Qy         110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2093 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2152

Qy         170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2153 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2212

Qy         230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
               |||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db        2213 TTCGAATTTCTGCCATTCATCCGCTTATTATCAATTATTCAGGCGTAGC-ACCAGGCGTT 2271

Qy         290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2272 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2331

Qy         350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
               |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        2332 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 2391

Qy         410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2392 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2451

Qy         470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2452 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2511

Qy         530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2512 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2571

Qy         590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2572 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2631

Qy         650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2632 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2691

Qy         710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
               |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        2692 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 2751

Qy         770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2752 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2811

Qy         830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2812 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2871

Qy         890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2872 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2931

Qy         950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2932 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 2991

Qy        1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2992 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 3051

Qy        1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3052 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 3111

Qy        1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        3112 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 3171

Qy        1190 GAAAAGGATCCG 1201
                ||| |  ||
Db        3172 CAAAGTGCGTCG 3183


RESULT 15
AAZ06434
ID   AAZ06434 standard; DNA; 5838 BP.
XX
```

```
AC   AAZ06434;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Plasmid pRZTL1, target DNA for Tn5 transposase assay.
XX
KW   transposase; modified form; wildtype; multimeric; OE termini; IE termini;
KW   outside end termini; inside end termini; plasmid; repeat sequence;
KW   mutation; cyclic; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   insertion_seq   1. .19
FT                   /*tag=  a
FT   CDS             77. .1267
FT                   /*tag=  b
FT                   /gene= "TetR"
FT                   /function= "effluxes tet"
FT                   /product= "inner membrane protein"
FT                   /phenotype= "Tetracycline resistance"
FT   CDS             2301. .2960
FT                   /*tag=  c
FT                   /gene= "ChlR"
FT                   /function= "acetylates chl"
FT                   /product= "chloramphenicol acetyltransferase"
FT                   /function= "acetylates chl"
FT                   /phenotype= "chloramphenicol resistance"
FT   insertion_seq   4564. .4582
FT                   /*tag=  d
FT   CDS             4715. .5530
FT                   /*tag=  e
FT                   /function= "blocks action of kanamycin"
FT                   /phenotype= "kanamycin resistance"
XX
PN   US5948622-A.
XX
PD   07-SEP-1999.
XX
PF   06-OCT-1997;   97US-00944916.
XX
PR   09-SEP-1996;   96US-00814877.
PR   02-MAY-1997;   97US-00850880.
XX
PA   (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
PI   Zhou H,  York DL,  Goryshin IY,  Reznikoff WS;
XX
DR   WPI; 1999-517947/43.
DR   P-PSDB; AAY15383, AAY15384, AAY15385.
XX
PT   In vitro transposition using a Tn5 based genetic construct.
XX
PS   Example 1; Col 31-39; 48pp; English.
XX
CC   This plasmid can be used in the assay of transposition frequency by the
CC   modified Tn5 transposase. The insertion of the Tn5 transposase releases a
CC   fragment of the plasmid that corresponds to the region of Kanamycin
CC   resistance gene that lies between the two OE sequences. The plasmid was
CC   also used to investigate the nature of the termini involved in the
CC   transposition reaction. Wildtype Outside End (OE,AAZ06435) and Inside End
CC   (IE, AAZ06438) were compared and an effort made to randomise the
CC   nucleotides at each of the seven positions of difference. This eventually
CC   lead to the identification of the mutant OE sequences AAZ06436 and
CC   AAZ06437
XX
SQ   Sequence 5838 BP; 1370 A; 1525 C; 1516 G; 1427 T; 0 U; 0 Other;

  Query Match             88.9%;  Score 1114.4;  DB 2;  Length 5838;
  Best Local Similarity   98.5%;  Pred. No. 0;
  Matches 1135;  Conservative    0;  Mismatches   16;  Indels    1;  Gaps    1;

Qy         50 AAAACGGATCCTATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 109
```

```
                ||  ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 ATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCA 2092

Qy        110 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2093 GGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCA 2152

Qy        170 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2153 ATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT 2212

Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              |||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||||
Db       2213 TTCGAATTTCTGCCATTCATCCGCTTATTATCAATTATTCAGGCGTAGC-ACCAGGCGTT 2271

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2272 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 2331

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       2332 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCATGATGAACC 2391

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2392 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 2451

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2452 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 2511

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2512 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 2571

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2572 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 2631

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2632 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 2691

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db       2692 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 2751

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2752 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 2811

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2812 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 2871

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2872 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 2931

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2932 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT 2991

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2992 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 3051

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 1129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3052 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACA 3111

Qy       1130 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGTC 1189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3112 GGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCG 3171

Qy       1190 GAAAAGGATCCG 1201
              |||  |   ||
Db       3172 CAAAGTGCGTCG 3183


Search completed: January 17, 2006, 19:35:10
Job time : 784.788 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 5307.56 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-2
Perfect score: 1254
Sequence: 1 gatctgattgattgaacgaa.....ttcgttcaatcaatcagatc 1254

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	867.4	69.2	971	10	CL111359	CL111359 ISB1-55D1
2	858.8	68.5	1039	10	CL094006	CL094006 ISB1-24F6
3	854.6	68.1	973	10	CL112191	CL112191 ISB1-56K1
4	852	67.9	968	10	CL131987	CL131987 ISB1-101G
5	852	67.9	994	10	CL134875	CL134875 ISB1-105K
6	842.4	67.2	1018	10	CL132765	CL132765 ISB1-102I
7	838	66.8	1047	10	CL091480	CL091480 ISB1-20I9
8	832.2	66.4	1001	10	CL145346	CL145346 ISB1-145L
9	831.4	66.3	971	10	CL129946	CL129946 ISB1-97H8
10	829.8	66.2	968	10	CL131806	CL131806 ISB1-101C
11	824.6	65.8	1007	10	CL112441	CL112441 ISB1-57C1
12	815.8	65.1	945	11	CT015391	CT015391 KBrH124J0
13	814	64.9	923	10	CL132302	CL132302 ISB1-101N
14	812.2	64.8	952	10	CZ941767	CZ941767 311863 To
15	803.6	64.1	901	10	CL129968	CL129968 ISB1-97H2
16	798.4	63.7	972	10	CL091218	CL091218 ISB1-20B5
17	797.4	63.6	939	9	CC219264	CC219264 CH261-114
18	792.6	63.2	1001	10	CL120945	CL120945 ISB1-79L1
19	783.2	62.5	1052	10	CL114391	CL114391 ISB1-60C1
20	781.4	62.3	1061	10	CL113434	CL113434 ISB1-58I1
21	780.8	62.3	831	9	BH239118	BH239118 ATYCK59TF
22	772	61.6	1078	10	CL110921	CL110921 ISB1-54H1
23	769.4	61.4	850	8	CV979827	CV979827 UMC-bemiv
24	769.2	61.3	1007	10	CL114392	CL114392 ISB1-60C1
25	767	61.2	965	10	CL134505	CL134505 ISB1-105B
26	765.8	61.1	812	10	DU053555	DU053555 141529 To
27	765.4	61.0	984	10	CL132776	CL132776 ISB1-102I
28	762	60.8	1004	10	CL039461	CL039461 CH216-47O
29	760.6	60.7	1011	10	CL111697	CL111697 ISB1-55K1
30	760.4	60.6	1026	10	CL114498	CL114498 ISB1-60F2
31	759.2	60.5	765	10	CL133473	CL133473 ISB1-103I
32	758.8	60.5	955	10	CL144387	CL144387 ISB1-144G
33	756.8	60.4	1207	9	CC279944	CC279944 CH261-24C
34	756.6	60.3	1027	10	CL111508	CL111508 ISB1-55G8
35	756.4	60.3	794	9	BH243026	BH243026 AUITG06TR
36	756.2	60.3	1010	10	CL112625	CL112625 ISB1-57G1
37	754.4	60.2	989	10	CL132140	CL132140 ISB1-101K
38	752.4	60.0	969	10	CL124573	CL124573 ISB1-85K6
39	750.2	59.8	782	9	BH242215	BH242215 ATZFD43TF
40	750.2	59.8	801	10	DU058617	DU058617 90433 Tom
41	750	59.8	934	10	CL124145	CL124145 ISB1-85A9
42	749.2	59.7	939	10	CL132916	CL132916 ISB1-102L
43	748.6	59.7	978	10	CL033185	CL033185 CH216-36C
44	748.2	59.7	985	10	CL131971	CL131971 ISB1-101G
45	746.8	59.6	946	10	CL145878	CL145878 ISB1-146H

ALIGNMENTS

RESULT 1
CL111359
LOCUS CL111359 971 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-55D1_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-55D1, genomic survey sequence.
ACCESSION CL111359
VERSION CL111359.1 GI:40604994
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 971)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 8
High quality sequence stop: 824.
FEATURES Location/Qualifiers
source 1. .971
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-55D1"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

Query Match 69.2%; Score 867.4; DB 10; Length 971;
Best Local Similarity 99.1%; Pred. No. 8.6e-254;
Matches 882; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy 231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 83 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 142

```
Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 202

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 262

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 322

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 382

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 442

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 502

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 562

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        563 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 622

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 682

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 742

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 802

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        803 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATA 862

Qy       1011 ACTCAAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTA 1070
              |||| |||||||||||||| |||||||||||||||||| ||||||||||||||||||||
Db        863 ACTC-AAAAATACGCCCGGGAGTGATCTTATTTCATTATGGNGAAAGTTGGAACCTCTTA 921

Qy       1071 CGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCG 1120
              || |||||| |||||||| |||||||||||||||||||||||||||||||
Db        922 CGGGCCGATTAACGTCTCAATTTCGCCAAAAGTTGGCCCAGGGCTTCCCG 971
```

```
RESULT 2
CL094006
LOCUS       CL094006                1039 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-24F6_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-24F6,
            genomic survey sequence.
ACCESSION   CL094006
VERSION     CL094006.1  GI:40587641
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1039)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 2
            High quality sequence stop: 744.
FEATURES             Location/Qualifiers
     source          1. .1039
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-24F6"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN
```

```
  Query Match             68.5%;  Score 858.8;  DB 10;  Length 1039;
  Best Local Similarity   96.4%;  Pred. No. 3.8e-251;
  Matches  889;  Conservative    0;  Mismatches   32;  Indels    1;  Gaps    1;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 126

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 186

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 246

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        247 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCTTATAATATTTGCCCATGGTGAAAA 306

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 366

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 426

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 486

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 546

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        547 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 606

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 666

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 726

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
```

```
Db        727 GAGCAACTGACTGAAATGCCTTCAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 786

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||| | |
Db        787 TGGTATATCCAGTGAATTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCCATA 846

Qy       1011 ACTCAAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTA 1070
              |||| |||||||| |||||||||||||||| ||| ||||||||||||||||||||||||||
Db        847 ACTC-AAAAATACCCCCGGTAGTGATCTTTTTTTATTATGGTGAAAGTTGGAACCTCTTA 905

Qy       1071 CGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAG 1130
              |||||| ||||||||||  ||||| |||||||| ||||  ||  | |||||||||| | |
Db        906 CGTGCCCATCAACGTCTATTTTTCCCCAAAAGTGGGCCAGGGCTTCCCCGGTATCACCGG 965

Qy       1131 GGACACCAGGATTTATTTATTC 1152
              |    || |  ||| |||  ||
Db        966 GACACCCGGATTTTCTTTTCTC 987


RESULT 3
CL112191
LOCUS       CL112191                973 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-56K13_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-56K13,
            genomic survey sequence.
ACCESSION   CL112191
VERSION     CL112191.1  GI:40605826
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 2
            High quality sequence stop: 769.
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-56K13"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match            68.1%;  Score 854.6;  DB 10;  Length 973;
  Best Local Similarity  98.9%;  Pred. No. 7.2e-250;
  Matches  871;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy        307 GCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAG 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAG 61

Qy        367 CATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCAT 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCAT 121

Qy        427 CAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTT 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTT 181

Qy        487 GTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGAC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGAC 241

Qy        547 GAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGC 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGC 301

Qy        607 CACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 CACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAG 361

Qy        667 CGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCA 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCA 421

Qy        727 TATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCATCAGGCG 786
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        422 TATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCG 481

Qy        787 GGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAA 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 GGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAA 541

Qy        847 AAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAA 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 AAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAA 601

Qy        907 TGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGAT 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 TGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGAT 661

Qy        967 TTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAATACGCC 1026
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        662 TTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCC 721

Qy       1027 CGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTC 1086
              |||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db        722 CGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATTCACGTC 781

Qy       1087 TCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGA-TTTA 1145
              |||||||||| |||  |||||||||||||||||||||||||||||||||||| ||||
Db        782 TCATTTTCGCCCAAAAGTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTTA 841

Qy       1146 TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCG 1186
              ||||||||||||||||||||||| | ||||||||||||
Db        842 TTTATTCTGCGAAGTGATCTTCCGTCCCCGGTATTTATTCG 882


RESULT 4
CL131987
LOCUS       CL131987                968 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-101G15_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-101G15,
            genomic survey sequence.
ACCESSION   CL131987
VERSION     CL131987.1  GI:40625622
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 968)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
```

```
          Class: BAC ends
          High quality sequence start: 16
          High quality sequence stop: 847.
FEATURES             Location/Qualifiers
     source          1. .968
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-101G15"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             67.9%;  Score 852;  DB 10;  Length 968;
  Best Local Similarity   99.2%;  Pred. No. 4.5e-249;
  Matches  877;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 145

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 205

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 265

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        266 GAATCGCCAGCGGCATCAGTACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 325

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 385

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 445

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 505

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 565

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db        566 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 625

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 685

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 745

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 805

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        806 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATA 865

Qy       1011 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTT-GGAACCTCTT 1069
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        866 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGGAACCTCTT 925

Qy       1070 ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGG 1113
              ||||| ||||| |||||||||||||||||||||||||||||| ||
Db        926 ACGTGGCGATC-ACGTCTCATTTTCGCCAAAAGTTGGCCCAAGG 968


RESULT 5
CL134875
LOCUS       CL134875                994 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-105K7_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-105K7,
            genomic survey sequence.
ACCESSION   CL134875
VERSION     CL134875.1  GI:40628510
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 6
            High quality sequence stop: 804.
FEATURES             Location/Qualifiers
     source          1. .994
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-105K7"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             67.9%;  Score 852;  DB 10;  Length 994;
  Best Local Similarity   97.5%;  Pred. No. 4.5e-249;
  Matches  887;  Conservative    0;  Mismatches   20;  Indels    3;  Gaps    2;

Qy        245 TTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAA 304
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          1 TTCTGCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCATTAA 60

Qy        305 CTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTA 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTA 120

Qy        365 AGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGC 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGC 180

Qy        425 ATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAG 484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAG 240

Qy        485 TTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAG 544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAG 300

Qy        545 ACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACAC 604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACAC 360

Qy        605 GCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAG 664
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAG 420

Qy        665 AGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCC 724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCC 480

Qy        725 CATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCATCAGG 784
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db        481 CATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGG 540

Qy        785 CGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTT 844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTT 600

Qy        845 AAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGA 904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGA 660

Qy        905 AATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTG 964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTG 720

Qy        965 ATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAATACG 1024
              | ||||||||||||||||||||||||||| ||||||||||| |||||||||||||||||
Db        721 AATTTTTTCTCCATTTTAGCTTCCTTAGCTCCCGAAAATCTCGATAACTCAAAAAATACG 780

Qy       1025 CCCGGTAGTGATCTTATTTCATTATGGTGAAAG-TTGGAACCTCTTACGTGCCGATCAAC 1083
              ||||||||||||||||||||||||||||||||| |||||||||||||| |||||||||
Db        781 CCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTTGGAACCTCTTACGGCCGATCAAC 840

Qy       1084 GTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGG--TATCAACAGGGACACCAGGA 1141
              |||||||||  ||||||||||||||  |||| ||||    |||||  |||||||||||
Db        841 GTCTCATTTTCCCAAAAGTTGGCCCAAAGCTTTCCGGGTAATCAAAGGGGACACCAGGA 900

Qy       1142 TTTATTTATT 1151
               ||  || ||
Db        901 ATTTATTTTT 910


RESULT 6
CL132765
LOCUS       CL132765               1018 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-102I9_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-102I9,
            genomic survey sequence.
ACCESSION   CL132765
VERSION     CL132765.1  GI:40626400
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1018)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 10
            High quality sequence stop: 781.
FEATURES             Location/Qualifiers
     source          1..1018
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-102I9"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             67.2%;  Score 842.4;  DB 10;  Length 1018;
  Best Local Similarity   95.7%;  Pred. No. 4e-246;
  Matches  898;  Conservative    0;  Mismatches   36;  Indels    4;  Gaps    3;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 144

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 204

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 264

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 324

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 384

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 444

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 504

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        505 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 564

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db        565 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 624

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 684

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db        685 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGGTATAGGTACATT 744

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        745 GAGCAACTGACTGAAATGCCTCCAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 804

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              ||||||||||| |||||||||||||||||||||||||| ||| |||||||||||||||| |
Db        805 TGGTATATCCAGGGATTTTTTTCTCCATTTTAGCTTCTTTAACTCCTGAAAATCTCGATA 864

Qy       1011 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTA 1070
              |||  ||||||| ||||||| |||||||||||||||| | |||||||||||||||||
Db        865 ACTTCAAAAAATACCCCCGGTAGGGATCTTATTTCATTATTGGGAAAGTTGGAACCTCTTA 924

Qy       1071 CGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAG 1130
              || ||| || ||||||| |||||| ||||| | |||| |||  || |||||||| ||||
Db        925 CGGGCCAAT-AACGTCTTATTTTCCCCAAAGTTGGGCCAAGG--TTTCCGGTATC-ACAG 980

Qy       1131 GGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCC 1168
              ||||||| ||| || ||| | | |  | |  | | ||
Db        981 GGACACCCGGAATTTTTTTTCCGGAAAGGGATTTTCCC 1018
```

```
RESULT 7
CL091480
LOCUS       CL091480               1047 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-20I9_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-20I9,
            genomic survey sequence.
ACCESSION   CL091480
VERSION     CL091480.1  GI:40585115
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1047)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 29
            High quality sequence stop: 769.
FEATURES             Location/Qualifiers
     source          1. .1047
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-20I9"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             66.8%;  Score 838;  DB 10;  Length 1047;
  Best Local Similarity   97.7%;  Pred. No. 9e-245;
  Matches  850;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        324 CCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGA 383
              ||||||||||||||||||  |  ||||||||||||||||||||||||||||||||||||
Db          1 CCCCGCCCTGCCACTCATCCCTTTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGA 60

Qy        384 AGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTT 443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTT 120

Qy        444 GCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGT 503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGT 180

Qy        504 TTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAA 563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAA 240

Qy        564 TAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA 623
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA 300

Qy        624 TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAG 683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAG 360

Qy        684 TTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGT 743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGT 420

Qy        744 CTTTCATTGCCATACGTAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAA 803
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAA 480

Qy        804 AGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCA 863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCA 540

Qy        864 GCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT 923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT 600

Qy        924 TACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAG 983
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAG 660

Qy        984 CTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTATTT 1043
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        661 CTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTT 720

Qy       1044 CATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGT 1103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGT 780

Qy       1104 TGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGAT 1163
              |||||||||||| |||||||||||||||||||||||||||||||||| || ||||
Db        781 TGGCCCAGGGCTCCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCCAAATGAT 840

Qy       1164 CTTCCGTCACAGGTATTTATTCGGTCGAAA 1193
              |||||||| |  | | ||  |    | |||
Db        841 CTTCCGTCCCCAGGAATTTATTCCCCAAAA 870


RESULT 8
CL145346
LOCUS       CL145346               1001 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-145L7_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-145L7,
            genomic survey sequence.
ACCESSION   CL145346
VERSION     CL145346.1  GI:40638981
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1001)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 10
            High quality sequence stop: 776.
FEATURES             Location/Qualifiers
     source          1. .1001
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-145L7"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             66.4%;  Score 832.2;  DB 10;  Length 1001;
  Best Local Similarity   96.9%;  Pred. No. 5.3e-243;
  Matches  890;  Conservative    0;  Mismatches   24;  Indels    4;  Gaps    4;
```

```
Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 146

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 206

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 266

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        267 GAATCGCCAGCGGCATCATTACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 326

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 386

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 446

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 506

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 566

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        567 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 626

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        627 GAGCATTCATCANGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 686

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        687 TCTTTACGGTC-TTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 745

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 805

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              |||||||||||||||||||||||| |||||||| |||||||| |||||||||||||| |
Db        806 TGGTATATCCAGTGATTTTTTTCCCCATTTTA-CTTCCTTAACTCCTGAAAATCTCGATA 864

Qy       1011 ACTCAAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTA 1070
              |||||||||| |||||||| ||||||||| |||    ||||| |||||| ||||||||||
Db        865 ACTCAAAAAA-ACGCCCGGGAGTGATCTTTTTTTCATATGGGGAAAGTGGGAACCTCTTA 923

Qy       1071 CGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAG 1130
              | |||||| ||||||| ||||| |||||||||||||||||||||||||| ||||| |
Db        924 GGGGCCGATAAACGTCT-ATTTTTCCCAAAAGTTGGCCCAGGGCTTCCCGGTTTCAACGG 982

Qy       1131 GGACACCAGGATTTATTT 1148
              |||| || |||||| |||
Db        983 GGACCCCGGGATTTTTTT 1000


RESULT 9
CL129946
LOCUS       CL129946                971 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-97H8_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-97H8,
            genomic survey sequence.
ACCESSION   CL129946
VERSION     CL129946.1  GI:40623581
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 10
            High quality sequence stop: 786.
FEATURES             Location/Qualifiers
     source          1. .971
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-97H8"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             66.3%;  Score 831.4;  DB 10;  Length 971;
  Best Local Similarity   97.3%;  Pred. No. 9.2e-243;
  Matches  877;  Conservative    0;  Mismatches   21;  Indels    3;  Gaps    3;

Qy        284 GGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCG 343
              ||  |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          3 GGTTTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCTTCG 62

Qy        344 CAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGA 403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGA 122

Qy        404 TGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATG 463
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATG 182

Qy        464 GTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAA 523
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAA 242

Qy        524 CTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAG 583
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTTTTTTTTTTTTTTAGGGAAATAG 302

Qy        584 GCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAA 643
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAA 362

Qy        644 TCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTG 703
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTG 422

Qy        704 TAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAAT 763
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        423 TAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAAT 482

Qy        764 TCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGC 823
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 TCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGC 542

Qy        824 TTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAG 883
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAG 602
```

```
Qy        884 GTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATA 943
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 GTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATA 662

Qy        944 TCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAAT 1003
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 TCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAAT 722

Qy       1004 CTCGACAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAA 1063
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 CTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAA 782

Qy       1064 CCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGC-TTCCCGGT 1122
              |||||||||||||||||||||||||||||  ||||||||||||||||||| ||||||||
Db        783 CCTCTTACGTGCCGATCAACGTCTCATTTTTCCCAAAAGTTGGCCCAGGGCTTTCCCGGT 842

Qy       1123 ATCAACAGGGACACCAGGA-TTTATTTATTCTGCGAAGTGAT-CTTCCGTCACAGGTATT 1180
              |||||||||||  ||||||||||||||||||||||||||| |||||||| ||||| ||
Db        843 ATCAACAGGGAACCCAGGATTTTATTTATTCTGCGAAGTGATCCTTCCGTCCCAGGTTTT 902

Qy       1181 T 1181
              |
Db        903 T 903
```

```
RESULT 10
CL131806
LOCUS       CL131806                968 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-101C13_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-101C13,
            genomic survey sequence.
ACCESSION   CL131806
VERSION     CL131806.1  GI:40625441
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 968)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 8
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .968
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-101C13"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             66.2%;  Score 829.8;  DB 10;  Length 968;
  Best Local Similarity   97.7%;  Pred. No. 2.9e-242;
  Matches  862;  Conservative    0;  Mismatches   18;  Indels    2;  Gaps    2;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 145

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 205

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 265

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        266 GAATCGCCAGCGGCATCATTACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 325

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 385

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 445

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 505

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 565

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        566 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 625

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 685

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 745

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||    ||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 GAGCAACTNGACTGAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 805

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||| |
Db        806 TGGTATATCCAGGGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATA 865

Qy       1011 ACTCAAAAAATACGCCC-GGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTT 1069
              |||||||||||||||| || || |||||||||||||||||| ||||||||||||||||||
Db        866 ACTCAAAAAATACGCCCGGGAAGGGATCTTATTTCATTATGGGGAAAGTTGGAACCTCTT 925

Qy       1070 ACGTG-CCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCA 1110
              ||||| |||||||||||||||||||  | |||||||| |||
Db        926 ACGTGCCCGATCAACGTCTCATTTTCTGCCAAAGTTGGGCCA 967
```

```
RESULT 11
CL112441
LOCUS       CL112441                1007 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-57C16_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-57C16,
            genomic survey sequence.
ACCESSION   CL112441
VERSION     CL112441.1  GI:40606076
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1007)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
```

```
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 8
            High quality sequence stop: 801.
FEATURES             Location/Qualifiers
     source          1. .1007
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-57C16"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN

  Query Match             65.8%;  Score 824.6;  DB 10;  Length 1007;
  Best Local Similarity   95.4%;  Pred. No. 1.1e-240;
  Matches  882;  Conservative    0;  Mismatches   39;  Indels    4;  Gaps    3;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 142

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 202

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 262

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 322

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 382

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 442

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 502

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 562

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        563 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 622

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 682

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 742

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 802

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTG-AAAATCTCGAC 1009
              ||||||||||||||||||||||||||||||||||||||||||||||| |||||||| |
Db        803 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAAATCTCAAT 862

Qy       1010 AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAA-CCTCT 1068
              ||||||||||     || || |||||||||| |||||||||| |||||||||| ||||| 
Db        863 AACTCAAAAATACGCCCGGGAAGTGATCTTAATTCATTATGGGGAAAGTTGGAACCCTCT 922

Qy       1069 TACGTGCCGATCAACGTCTC--ATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCA 1126
              |||| |||||| || |||||    ||| | | |||  ||| ||||||||||||| | | 
Db        923 TACGGGCCGATTAAAGTCTCCATTTTCCCCAAAAGTTTGCCCCAGGGCTTCCCCGGTTTC 982

Qy       1127 ACAGGGACACCAGGATTTATTTATT 1151
              | | ||  |||      | ||| ||
Db        983 AAACGGGAACCCCCGGATTTTTTTT 1007


RESULT 12
CT015391
LOCUS       CT015391                945 bp    DNA     linear   GSS 09-AUG-2005
DEFINITION  KBrH124J09 genomic clone, KBrH (HindIII) BAC library Brassica rapa
            subsp. pekinensis, genomic survey sequence.
ACCESSION   CT015391
VERSION     CT015391.1  GI:71467442
KEYWORDS    GSS.
SOURCE      Brassica rapa subsp. pekinensis
  ORGANISM  Brassica rapa subsp. pekinensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1
  AUTHORS   Viehoever,P., Holtgraewe,D. and Weisshaar,B.
  TITLE     BAC end sequences of Brassica rapa
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 945)
  AUTHORS   Li,Y. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2005) Weisshaar B., Bielefeld University,
            Institute for Genome Research, Universitaetsstrasse 25, D-33594
            Bielefeld, Germany
COMMENT     Contact: Bernd Weisshaar
            Bielefeld University, Institute for Genome Research
            Universitaetsstrasse 25, D-33594 Bielefeld, Germany Email:
            bernd.weisshaar@uni-bielefeld.de
            BAC end sequences of Brassica rapa BAC clone KBrH124J09; generated
            as contribution to the 'Multinational Brassica rapa Sequencing
            Project' Seq primer: m13f TGTAAAACGACGGCCAGT
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Brassica rapa subsp. pekinensis"
                     /mol_type="genomic DNA"
                     /strain="Chiifu type 401-42"
                     /cultivar="Chiifu"
                     /sub_species="pekinensis"
                     /db_xref="taxon:51351"
                     /clone="KBrH124J09"
                     /clone_lib="KBrH, Brassica rapa HindIII BAC library
                     GF-SCF-1002, Vector: pCUGIBac1"
                     /lab_host="E.coli DH10B"
ORIGIN

  Query Match             65.1%;  Score 815.8;  DB 11;  Length 945;
  Best Local Similarity   99.8%;  Pred. No. 5.5e-238;
  Matches  817;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 186

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 246
```

```
Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 306

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 366

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 426

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 486

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 546

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 606

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        607 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGAT 666

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        667 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 726

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 786

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 846

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        847 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATA 906

Qy       1011 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTAT 1049
              |||||||||||||||||||||||||||||||||||||||
Db        907 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTAT 945
```

```
RESULT 13
CL132302
LOCUS       CL132302                923 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-101N19_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-101N19,
            genomic survey sequence.
ACCESSION   CL132302
VERSION     CL132302.1  GI:40625937
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 12
            High quality sequence stop: 877.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-101N19"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
                     Library Segment 1"
ORIGIN
```

```
  Query Match             64.9%;  Score 814;  DB 10;  Length 923;
  Best Local Similarity   99.3%;  Pred. No. 2e-237;
  Matches  828;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy        231 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         89 TCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTT 148

Qy        291 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        149 AAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACT 208

Qy        351 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 GTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCT 268

Qy        411 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 GAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAA 328

Qy        471 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 CGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCC 388

Qy        531 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        389 AGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGT 448

Qy        591 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        449 TTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGT 508

Qy        651 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        509 GGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAG 568

Qy        711 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGAT 770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        569 GGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT 628

Qy        771 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 830
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        629 GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT 688

Qy        831 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 890
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        689 TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATT 748

Qy        891 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 950
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        749 GAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGG 808

Qy        951 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACA 1010
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        809 TGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCCTGAAATCTCGATA 868

Qy       1011 ACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAAC 1064
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        869 ACTC-AAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAAC 921


RESULT 14
CZ941767
LOCUS       CZ941767                952 bp    DNA     linear   GSS 11-AUG-2005
DEFINITION  311863 Tomato EcoRI BAC Library Lycopersicon esculentum genomic
            clone SL_EcoRI0034N17 5, genomic survey sequence.
ACCESSION   CZ941767
VERSION     CZ941767.1  GI:72276159
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 952)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 254990
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 34  row: N  column: 17
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 19
            High quality sequence stop: 680.
FEATURES             Location/Qualifiers
     source          1. .952
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="SL_EcoRI0034N17"
                     /lab_host="E. coli"
                     /clone_lib="Tomato EcoRI BAC Library"
                     /note="Vector: unk; Site_1: EcoRI"
ORIGIN

  Query Match             64.8%;  Score 812.2;  DB 10;  Length 952;
  Best Local Similarity   97.6%;  Pred. No. 7e-237;
  Matches  856;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    3;

Qy        309 CTTAAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCA 368
              | | |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         19 CCTTAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGCAATTCATTAAGCA 78

Qy        369 TTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCA 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 TTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCA 138

Qy        429 GCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGT 488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 GCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGT 198

Qy        489 CCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGA 548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 CCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGA 258

Qy        549 AAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCA 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 AAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCA 318

Qy        609 CATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCG 668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 CATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCG 378

Qy        669 ATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA 728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 ATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATA 438

Qy        729 TCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGATGAGCATTCATCAGGCGGG 788
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 TCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGG 498

Qy        789 CAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAA 848
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        499 CAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAA 558

Qy        849 AGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATG 908
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 AGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATG 618

Qy        909 CCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT 968
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        619 CCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT 678

Qy        969 TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGACAACTCAAAAAATACGCCCG 1028
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        679 TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCG 738

Qy       1029 GTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTC 1088
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        739 GTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCCTACGTGCCGATCAACGTCTC 798

Qy       1089 ATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTT 1148
              |||||||||||||| ||||||  |||||||| ||||||||||||||||| ||||||||||
Db        799 ATTTTCGCCAAAAG-TGGCCCCAGGCTTCCCCGTATCAACAGGGACACC-GGATTTATTT 856

Qy       1149 ATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTC 1185
              || |||| |||||||||||||||  ||   ||  | |
Db        857 ATCCTGC-AAGTGATCTTCCGTCCAAGTATTTATTCC 892


RESULT 15
CL129968
LOCUS       CL129968                901 bp    DNA     linear   GSS 05-JAN-2004
DEFINITION  ISB1-97H21_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-97H21,
            genomic survey sequence.
ACCESSION   CL129968
VERSION     CL129968.1  GI:40623603
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 75000   Std Error: 0.00
            Seq primer: T7 TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 18
            High quality sequence stop: 734.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:8364"
                     /clone="ISB1-97H21"
                     /clone_lib="ISB1"
                     /note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC
```

Library Segment 1"

ORIGIN

Query Match 64.1%; Score 803.6; DB 10; Length 901;
Best Local Similarity 96.7%; Pred. No. 3e-234;
Matches 853; Conservative 0; Mismatches 24; Indels 5; Gaps 3;

```
Qy        230 TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 289
              | |||||||||| | ||||||||||||||||||||||||||||||||||||||||||||
Db          3 TCCGAATTTCTGGCCTTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTT 62

Qy        290 TAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 349
              |   |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TCTCGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC 122

Qy        350 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACC 182

Qy        410 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 TGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAA 242

Qy        470 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 ACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC 302

Qy        530 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGG 362

Qy        590 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 649
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG 422

Qy        650 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 709
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAA 482

Qy        710 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGTAATTCCGGA 769
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        483 GGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGA 542

Qy        770 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTT 602

Qy        830 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACAT 662

Qy        890 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 949
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        663 TGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACG 722

Qy        950 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAC 1009
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        723 GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTTCTGAAAATCTCGAT 782

Qy       1010 AACTCAAAAAATACGCCC-GGTAGTGATC-TTATTTCATTATGGTGAAAGTTGGAACCTC 1067
              |||||||||||||||||| |||||||||| || ||||||||||||||||||||||||||
Db        783 AACTCAAAAAATACGCCCGGGTAGTGATCTTTTTTTCATTATGGTGAAAGTTGGAACCTC 842

Qy       1068 TT---ACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGG 1106
              ||     |  |  || |||||||| ||||  || |||  | |
Db        843 TTTACGTGCCCGAATAAACGTCTCTTTTTTCCCCAAAAGTTG 884
```

Search completed: January 18, 2006, 11:40:49
Job time : 5311.56 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 257.623 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-5
Perfect score: 50
Sequence: 1 tgattgattgaacgaaaaac.....ttcacgataaatgcgaaaac 50

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	39.4	78.8	58	6	AR084420	AR084420 Sequence
c	2	39.4	78.8	71	6	AR084419	AR084419 Sequence
	3	39.4	78.8	77	6	BD073246	BD073246 In vitro
	4	39.4	78.8	77	6	AR353866	AR353866 Sequence
c	5	39.4	78.8	82	7	PMUBSR	M10177 Bacteriopha
c	6	39.4	78.8	83	6	AR084418	AR084418 Sequence
c	7	39.4	78.8	117	6	AR084417	AR084417 Sequence
	8	39.4	78.8	120	11	SYNGENE	M12560 Synthetic g
	9	39.4	78.8	150	7	PMUSEE1	M10863 Bacteriopha
c	10	39.4	78.8	220	7	NCMU3R	X05582 Mu-derived
c	11	39.4	78.8	220	7	PMUNE2	M34920 Bacteriopha
c	12	39.4	78.8	240	1	ECOTRPAA	M33723 Escherichia
c	13	39.4	78.8	324	11	SYNPMUEND	M15949 Synthetic B
	14	39.4	78.8	903	6	A02708	A02708 pNM506 DNA
	15	39.4	78.8	1319	11	AY781404	AY781404 Synthetic
c	16	39.4	78.8	1319	11	AY781404	AY781404 Synthetic
	17	39.4	78.8	1351	11	AY781405	AY781405 Synthetic
c	18	39.4	78.8	1351	11	AY781405	AY781405 Synthetic
	19	39.4	78.8	1524	11	AY781401	AY781401 Synthetic
c	20	39.4	78.8	1524	11	AY781401	AY781401 Synthetic
	21	39.4	78.8	1808	11	AY781408	AY781408 Synthetic
c	22	39.4	78.8	1808	11	AY781408	AY781408 Synthetic
	23	39.4	78.8	1937	11	AY781403	AY781403 Synthetic
c	24	39.4	78.8	1937	11	AY781403	AY781403 Synthetic
	25	39.4	78.8	2037	11	AY781407	AY781407 Synthetic
c	26	39.4	78.8	2037	11	AY781407	AY781407 Synthetic
	27	39.4	78.8	2062	11	AY781402	AY781402 Synthetic
c	28	39.4	78.8	2062	11	AY781402	AY781402 Synthetic
	29	39.4	78.8	2162	11	AY781406	AY781406 Synthetic
c	30	39.4	78.8	2162	11	AY781406	AY781406 Synthetic
	31	39.4	78.8	2413	11	AY781399	AY781399 Synthetic
c	32	39.4	78.8	2413	11	AY781399	AY781399 Synthetic
	33	39.4	78.8	2933	11	AY781400	AY781400 Synthetic
c	34	39.4	78.8	2933	11	AY781400	AY781400 Synthetic
	35	39.4	78.8	4666	1	ECOPOLBDA	M35371 E.coli DNA
	36	39.4	78.8	4791	11	AY738638	AY738638 Cloning v
c	37	39.4	78.8	4791	11	AY738638	AY738638 Cloning v
c	38	39.4	78.8	7003	6	AR084426	AR084426 Sequence
	39	39.4	78.8	15611	1	ECOPHNAQ	J05260 E.coli psiD
	40	39.4	78.8	15611	6	AR229537	AR229537 Sequence
c	41	39.4	78.8	36717	7	AF083977	AF083977 Bacteriop
c	42	39.4	78.8	37199	11	AY860420	AY860420 Cloning v
	43	37.4	74.8	240	1	ECOTRPAA	M33723 Escherichia
	44	31.8	63.6	5755	1	ECOCYS	M32101 E.coli thio
c	45	30.4	60.8	1925	1	ECOMALIXA	M28539 E.coli malI

ALIGNMENTS

RESULT 1
AR084420/c
LOCUS AR084420 58 bp DNA linear PAT 01-SEP-2000
DEFINITION Sequence 6 from patent US 5981177.
ACCESSION AR084420
VERSION AR084420.1 GI:10011191
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 58)
AUTHORS Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.
TITLE Protein fusion method and constructs
JOURNAL Patent: US 5981177-A 6 09-NOV-1999;
FEATURES Location/Qualifiers
source 1. .58
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match 78.8%; Score 39.4; DB 6; Length 58;
Best Local Similarity 97.6%; Pred. No. 0.0051;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
| ||||||||||||||||||||||||||||||||||||||||||
Db 49 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 9

RESULT 2
AR084419/c
LOCUS AR084419 71 bp DNA linear PAT 01-SEP-2000
DEFINITION Sequence 5 from patent US 5981177.
ACCESSION AR084419
VERSION AR084419.1 GI:10011190
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 71)
AUTHORS Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.

```
  TITLE     Protein fusion method and constructs
  JOURNAL   Patent: US 5981177-A 5 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .71
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 6;  Length 71;
  Best Local Similarity   97.6%;  Pred. No. 0.005;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         62 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 22


RESULT 3
BD073246
LOCUS       BD073246                 77 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  In vitro method for providing template for determination of DNA
            sequence.
ACCESSION   BD073246
VERSION     BD073246.1  GI:22618849
KEYWORDS    JP 2001510055-A/1.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 77)
  AUTHORS   Tenkanen,T., Soininen,T., Sabirarty,H. and Murtanen,K.
  TITLE     In vitro method for providing template for determination of DNA
  JOURNAL   Patent: JP 2001510055-A 1 31-JUL-2001;
            FINNZYMES OY
COMMENT     OS   Artificial Sequence
            PN   JP 2001510055-A/1
            PD   31-JUL-2001
            PF   10-JUL-1998 JP 2000503240
            PR   14-JUL-1997 FI       972992
            PI   THOMAS TENKANEN,TIMO SOININEN,HARI SABIRARTY,KIRSI MURTANEN PC
              C12Q1/68,C12N15/09,C12N15/00
            CC   Description of Artificial Sequence: transposon DNA capable of
            CC           forming a
            CC   functional transposon complex with MuA transposase FH   Key
                     Location/Qualifiers
            FT   source          1. .77
            FT                   /organism='Artificial Sequence'.
FEATURES             Location/Qualifiers
     source          1. .77
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 6;  Length 77;
  Best Local Similarity   97.6%;  Pred. No. 0.005;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         10 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50


RESULT 4
AR353866
LOCUS       AR353866                 77 bp    DNA     linear   PAT 17-AUG-2003
DEFINITION  Sequence 1 from patent US 6593113.
ACCESSION   AR353866
VERSION     AR353866.1  GI:33759950
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 77)
  AUTHORS   Tenkanen,T., Soininen,T., Savilahti,H. and Multanen,K.
  TITLE     In vitro method for providing templates for DNA sequencing
  JOURNAL   Patent: US 6593113-A 1 15-JUL-2003;
            Finnzymes Oy; Espoo;
            WOX;
FEATURES             Location/Qualifiers
     source          1. .77
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 6;  Length 77;
  Best Local Similarity   97.6%;  Pred. No. 0.005;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         10 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50


RESULT 5
PMUBSR/c
LOCUS       PMUBSR                   82 bp    DNA     linear   PHG 28-APR-1993
DEFINITION  Bacteriophage Mu right end with Mu A protein binding site R3.
ACCESSION   M10177
VERSION     M10177.1  GI:215519
KEYWORDS    .
SOURCE      Enterobacteria phage Mu
  ORGANISM  Enterobacteria phage Mu
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
            Mu-like viruses.
REFERENCE   1  (bases 1 to 30)
  AUTHORS   Craigie,R., Mizuuchi,M. and Mizuuchi,K.
  TITLE     Site-specific recognition of the bacteriophage Mu ends by the Mu A
            protein
  JOURNAL   Cell 39 (2 Pt 1), 387-394 (1984)
   PUBMED   6094016
REFERENCE   2  (bases 1 to 82)
  AUTHORS   Craigie,R.
  JOURNAL   Unpublished (1985)
COMMENT     Original source text: Bacteriophage Mu DNA.
            Computer-readable sequence for [1] kindly provided by R.Craigie,
            15-NOV-1985.
FEATURES             Location/Qualifiers
     source          1. .82
                     /organism="Enterobacteria phage Mu"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10677"
ORIGIN      Unreported.

  Query Match             78.8%;  Score 39.4;  DB 7;  Length 82;
  Best Local Similarity   97.6%;  Pred. No. 0.005;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         73 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 33


RESULT 6
AR084418/c
LOCUS       AR084418                 83 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 4 from patent US 5981177.
ACCESSION   AR084418
VERSION     AR084418.1  GI:10011189
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 83)
  AUTHORS   Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.
  TITLE     Protein fusion method and constructs
  JOURNAL   Patent: US 5981177-A 4 09-NOV-1999;
```

```
FEATURES             Location/Qualifiers
     source          1. .83
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 6;  Length 83;
  Best Local Similarity   97.6%;  Pred. No. 0.0049;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         74 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 34


RESULT 7
AR084417/c
LOCUS       AR084417                 117 bp    DNA     linear   PAT 01-SEP-2000
DEFINITION  Sequence 3 from patent US 5981177.
ACCESSION   AR084417
VERSION     AR084417.1  GI:10011188
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 117)
  AUTHORS   Demirjian,D.C., Casadaban,M.J., Weber,J.Mark. and Gaines,G.L. III.
  TITLE     Protein fusion method and constructs
  JOURNAL   Patent: US 5981177-A 3 09-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .117
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 6;  Length 117;
  Best Local Similarity   97.6%;  Pred. No. 0.0048;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db        108 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 68


RESULT 8
SYNGENE
LOCUS       SYNGENE                  120 bp    DNA     linear   SYN 27-APR-1993
DEFINITION  Synthetic gene utilizing phosphorylated synthetic fragments in
            vivo.
ACCESSION   M12560
VERSION     M12560.1  GI:208423
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 120)
  AUTHORS   Narang,S.A., Dubuc,G., Yao,F.L. and Michniewicz,J.J.
  TITLE     'In vitro' method of assembling a synthetic gene
  JOURNAL   Biochem. Biophys. Res. Commun. 134 (1), 407-411 (1986)
   PUBMED   3004442
COMMENT     Original source text: Phophorylated synthetic DNA fragments.
FEATURES             Location/Qualifiers
     source          1. .120
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN      EcoRI site.

  Query Match             78.8%;  Score 39.4;  DB 11;  Length 120;
  Best Local Similarity   97.6%;  Pred. No. 0.0048;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         15 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 9
PMUSEE1
LOCUS       PMUSEE1                  150 bp    DNA     linear   PHG 28-APR-1993
DEFINITION  Bacteriophage Mu DNA, SE end fragment.
ACCESSION   M10863
VERSION     M10863.1  GI:215593
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Enterobacteria phage Mu
  ORGANISM  Enterobacteria phage Mu
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
            Mu-like viruses.
REFERENCE   1  (bases 1 to 150)
  AUTHORS   Allet,B.
  TITLE     Nucleotide sequences at the ends of bacteriophage Mu DNA
  JOURNAL   Nature 274 (5671), 553-558 (1978)
   PUBMED   672985
COMMENT     Original source text: Bacteriophage Mu DNA.
FEATURES             Location/Qualifiers
     source          1. .150
                     /organism="Enterobacteria phage Mu"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10677"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 7;  Length 150;
  Best Local Similarity   97.6%;  Pred. No. 0.0048;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         18 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 58


RESULT 10
NCMU3R/c
LOCUS       NCMU3R                   220 bp    DNA     linear   PHG 02-APR-1988
DEFINITION  Mu-derived phage lambda placMu3 provirus right end (MuR) inserted
            in pBRG1214.
ACCESSION   X05582
VERSION     X05582.1  GI:15446
KEYWORDS    .
SOURCE      Enterobacteria phage Mu
  ORGANISM  Enterobacteria phage Mu
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
            Mu-like viruses.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   Nag,D.K. and Berg,D.E.
  TITLE     Specificity of bacteriophage Mu excision
  JOURNAL   Mol. Gen. Genet. 207 (2-3), 395-401 (1987)
   PUBMED   3039296
COMMENT     see x05580 for pBRG1214  with insertion sites
            see X05581 for prophage left end.
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Enterobacteria phage Mu"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10677"
     misc_feature    1. .220
                     /note="right end"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 7;  Length 220;
  Best Local Similarity   97.6%;  Pred. No. 0.0046;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db        211 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 171
```

```
RESULT 11
PMUNE2/c
LOCUS       PMUNE2                   220 bp    DNA     linear   PHG 28-APR-1993
DEFINITION  Bacteriophage mu genomic right end.
ACCESSION   M34920
VERSION     M34920.1  GI:215584
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Enterobacteria phage Mu
  ORGANISM  Enterobacteria phage Mu
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
            Mu-like viruses.
REFERENCE   1  (bases 1 to 220)
  AUTHORS   Groenen,M.A., Timmers,E. and van de Putte,P.
  TITLE     DNA sequences at the ends of the genome of bacteriophage Mu
            essential for transposition
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 82 (7), 2087-2091 (1985)
   PUBMED   2984681
COMMENT     Original source text: Bacteriophage mu DNA.
FEATURES             Location/Qualifiers
     source          1. .220
                     /organism="Enterobacteria phage Mu"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10677"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 7;  Length 220;
  Best Local Similarity   97.6%;  Pred. No. 0.0046;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db        211 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 171


RESULT 12
ECOTRPAA/c
LOCUS       ECOTRPAA                 240 bp    DNA     linear   BCT 21-JUN-2002
DEFINITION  Escherichia coli transposon Mu dl-R insertion site.
ACCESSION   M33723
VERSION     M33723.1  GI:208387
KEYWORDS    trp operon.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Chang,C.N., Kuang,W.J. and Chen,E.Y.
  TITLE     Nucleotide sequence of the alkaline phosphatase gene of Escherichia
            coli
  JOURNAL   Gene 44 (1), 121-125 (1986)
   PUBMED   3533724
REFERENCE   2  (bases 1 to 240)
  AUTHORS   Metcalf,W.W., Steed,P.M. and Wanner,B.L.
  TITLE     Identification of phosphate starvation-inducible genes in
            Escherichia coli K-12 by DNA sequence analysis of psi::lacZ(Mu dl)
            transcriptional fusions
  JOURNAL   J. Bacteriol. 172 (6), 3191-3200 (1990)
   PUBMED   2160940
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
     repeat_region   1. .104
                     /citation=[2]
                     /transposon="Mu dl-R"
     CDS             <105. .>240
                     /note="trp operon"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA75351.1"
                     /db_xref="GI:984720"
                     /translation="LRKGTMTTLLNPYFGEFGGMYVPQILMPALRQLEEAFVSAQKDP
                     E"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 1;  Length 240;
  Best Local Similarity   97.6%;  Pred. No. 0.0046;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         95 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 13
SYNPMUEND/c
LOCUS       SYNPMUEND                324 bp    DNA     linear   SYN 27-APR-1993
DEFINITION  Synthetic Bacteriophage Mu left and right end DNA.
ACCESSION   M15949
VERSION     M15949.1  GI:209098
KEYWORDS    synthetic sequence.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Patterson,T.A., Court,D.L., Dubuc,G., Michniewicz,J.J.,
            Goodchild,J., Bukhari,A.I. and Narang,S.A.
  TITLE     Transposition studies of mini-Mu plasmids constructed from the
            chemically synthesized ends of bacteriophage Mu
  JOURNAL   Gene 50 (1-3), 101-109 (1986)
   PUBMED   3034727
COMMENT     Original source text: Bacteriophage Mu DNA.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 11;  Length 324;
  Best Local Similarity   97.6%;  Pred. No. 0.0045;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db        314 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 274


RESULT 14
A02708
LOCUS       A02708                   903 bp    DNA     linear   PAT 27-APR-1993
DEFINITION  pNM506 DNA sequence.
ACCESSION   A02708
VERSION     A02708.1  GI:344652
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 903)
  AUTHORS   .
  TITLE     EXPRESSION VECTOR FOR ADJUSTABLE EXPRESSION OF EXOGENOUS GENES IN
            PROKARYOTES
  JOURNAL   Patent: WO 8809373-A 6 01-DEC-1988;
FEATURES             Location/Qualifiers
     source          1. .903
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     CDS             705. .788
                     /note="unnamed protein product; pNM506 mgl-B sequenz"
                     /codon_start=1
                     /transl_table=11
```

```
                     /protein_id="CAA00263.1"
                     /db_xref="GI:344653"
                     /translation="MNKKVLTLSAVMASLLFGAHAHAADTR"
ORIGIN

  Query Match            78.8%;  Score 39.4;  DB 6;  Length 903;
  Best Local Similarity  97.6%;  Pred. No. 0.0042;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db        795 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 835


RESULT 15
AY781404
LOCUS       AY781404               1319 bp    DNA     linear   SYN 28-FEB-2005
DEFINITION  Synthetic construct transposon mini-Mu transposon TnCR6, complete
            sequence.
ACCESSION   AY781404
VERSION     AY781404.1  GI:60171997
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 1319)
  AUTHORS   Zhang,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2004) CopyRat Pty Ltd, 27-31 Wright Street,
            Clayton, Victoria 3168, Australia
FEATURES             Location/Qualifiers
     source          1. .1319
                     /organism="synthetic construct"
                     /mol_type="other DNA"
                     /db_xref="taxon:32630"
     repeat_region   1. .1319
                     /transposon="mini-Mu transposon TnCR6"
     CDS             300. .959
                     /codon_start=1
                     /transl_table=11
                     /product="chloramphenicol acetyltransferase"
                     /protein_id="AAX14471.1"
                     /db_xref="GI:60171998"
                     /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                     LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                     SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                     ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                     A"
ORIGIN

  Query Match            78.8%;  Score 39.4;  DB 11;  Length 1319;
  Best Local Similarity  97.6%;  Pred. No. 0.0041;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         15 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


Search completed: January 18, 2006, 04:25:17
Job time : 258.623 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 31.092 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-5
Perfect score: 50
Sequence: 1 tgattgattgaacgaaaaac.....ttcacgataaatgcgaaaac 50

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	50	100.0	50	10	ACF58172	Acf58172 Cat-Mu tr
	2	50	100.0	54	10	ACF58168	Acf58168 Cat-Mu tr
	3	50	100.0	1254	10	ACF58169	Acf58169 Modified
c	4	50	100.0	1254	10	ACF58169	Acf58169 Modified
c	5	39.4	78.8	51	13	ADS13821	Ads13821 Mu end DN
	6	39.4	78.8	54	4	AAD21279	Aad21279 Precut tr
	7	39.4	78.8	54	10	ACF58171	Acf58171 Cat-Mu(No
	8	39.4	78.8	54	10	ACF58170	Acf58170 Cat-Mu tr
	9	39.4	78.8	56	13	ADS13820	Ads13820 Mu end DN
c	10	39.4	78.8	58	2	AAZ28883	Aaz28883 Mini-Mu t
c	11	39.4	78.8	71	2	AAZ28882	Aaz28882 Mini-Mu t
	12	39.4	78.8	82	10	ACC80745	Acc80745 Transposo
c	13	39.4	78.8	83	2	AAZ28881	Aaz28881 Mini-Mu t
	14	39.4	78.8	83	10	ACC80741	Acc80741 Transposo
	15	39.4	78.8	84	10	ACC80738	Acc80738 Transposo
	16	39.4	78.8	86	10	ACC80742	Acc80742 Transposo
	17	39.4	78.8	89	10	ACC80753	Acc80753 Transposo
	18	39.4	78.8	90	13	ADR14827	Adr14827 PCR prime
c	19	39.4	78.8	117	2	AAZ28880	Aaz28880 Mini-Mu t
	20	39.4	78.8	903	1	AAN80102	Aan80102 EcoRI/Bam
c	21	39.4	78.8	7003	2	AAZ28877	Aaz28877 Transposa
	22	39.4	78.8	15611	3	AAD01008	Aad01008 Escherich
	23	28	56.0	30	4	AAD21278	Aad21278 Bacteriop
c	24	25	50.0	34	14	AEB45577	Aeb45577 SigA2 wit
c	25	25	50.0	34	14	AEB48763	Aeb48763 Beta-lact
c	26	24.6	49.2	1874	6	AAS98204	Aas98204 DNA encod
c	27	24.6	49.2	2229	7	ADS73196	Ads73196 Human kid
c	28	24.6	49.2	2229	7	ADW42050	Adw42050 cDNA elev
c	29	24.6	49.2	6730	11	ADL22564	Adl22564 Human dis
c	30	24.6	49.2	17687	4	AAK71665	Aak71665 Human imm
c	31	24.6	49.2	17687	4	AAK64966	Aak64966 Human imm
c	32	24.6	49.2	17979	4	AAK71664	Aak71664 Human imm
c	33	24.6	49.2	17979	4	AAK64964	Aak64964 Human imm
c	34	24	48.0	2000	11	ACL38625	Acl38625 Rice stre
c	35	24	48.0	2099	6	ABQ69021	Abq69021 Listeria
c	36	24	48.0	7868	6	ABQ71053	Abq71053 Listeria
c	37	24	48.0	110000	6	ABQ69245_09	Continuation (10 o
c	38	24	48.0	110000	6	ABQ67197_08	Continuation (9 of
	39	23.8	47.6	2022	13	ADR85671	Adr85671 Aspergill
	40	23.8	47.6	2022	13	ADR85084	Adr85084 Aspergill
	41	23.8	47.6	8022	13	ADR84497	Adr84497 Aspergill
	42	23.6	47.2	917	6	ABQ25453	Abq25453 Oligonucl
c	43	23.6	47.2	917	6	ABQ25452	Abq25452 Oligonucl
c	44	23.6	47.2	955	6	ABQ32590	Abq32590 Oligonucl
	45	23.6	47.2	955	6	ABQ32591	Abq32591 Oligonucl

ALIGNMENTS

RESULT 1
ACF58172
ID ACF58172 standard; DNA; 50 BP.
XX
AC ACF58172;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cat-Mu transposon modified end fragment.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Claim 6; Page 31; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a

```
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a Cat-Mu transposon modified end
CC   fragment without 5' overhang
XX
SQ   Sequence 50 BP; 21 A; 8 C; 11 G; 10 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 50;  DB 10;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-09;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50


RESULT 2
ACF58168
ID   ACF58168 standard; DNA; 54 BP.
XX
AC   ACF58168;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Cat-Mu transposon modified end fragment.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Synthetic.
OS   Bacteriophage mu.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .4
FT                   /*tag=  a
FT                   /note= "the 5' end of this strand overhangs the 3'end of
FT                   the complementary strand"
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Claim 6; Fig 2; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a Cat-Mu transposon modified end
CC   fragment
XX
SQ   Sequence 54 BP; 22 A; 9 C; 12 G; 11 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 50;  DB 10;  Length 54;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-09;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 3
ACF58169
ID   ACF58169 standard; DNA; 1254 BP.
XX
AC   ACF58169;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Modified Cat-Mu(Stop)-transposon.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Synthetic.
OS   Bacteriophage mu.
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Claim 6; Page 30; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
SQ   Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 50;  DB 10;  Length 1254;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-08;
  Matches   50;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 4
ACF58169/c
ID   ACF58169 standard; DNA; 1254 BP.
```

XX
AC ACF58169;
XX
DT 15-JAN-2004 (first entry)
XX
DE Modified Cat-Mu(Stop)-transposon.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS Synthetic.
OS Bacteriophage mu.
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Claim 6; Page 30; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a
CC transposon containing a genetically engineered translation stop signal
CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a modified Cat-Mu(Stop)-transposon
XX
SQ Sequence 1254 BP; 356 A; 296 C; 264 G; 338 T; 0 U; 0 Other;

Query Match 100.0%; Score 50; DB 10; Length 1254;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1250 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 1201

RESULT 5
ADS13821/c
ID ADS13821 standard; DNA; 51 BP.
XX
AC ADS13821;
XX
DT 16-DEC-2004 (first entry)
XX
DE Mu end DNA fragment construction oligonucleotide MM1141.
XX
KW ss; mismatch detection; transposition detection; pathogen typing;
KW embryo screening; mutation detection; Mu end DNA; MM1141.
XX
OS Enterobacteria phage Mu.
OS Synthetic.
XX
PN US2004191821-A1.
XX
PD 30-SEP-2004.
XX
PF 26-MAR-2004; 2004US-00809688.
XX
PR 28-MAR-2003; 2003US-0457934P.
XX
PA (USSA) US SEC OF ARMY.
XX
PI Yanagihara K, Mizuuchi K;
XX
DR WPI; 2004-689846/67.
XX
PT Detecting a mismatch in a test double stranded nucleic acid target,
PT useful for typing a pathogenic microorganism strain, comprises detecting
PT transposition of the Mu-end nucleic acid into the target.
XX
PS Example 1; SEQ ID NO 2; 24pp; English.
XX
CC The invention relates to a method of detecting a mismatch in a test
CC double stranded nucleic acid target which comprises detecting
CC transposition of the Mu-end nucleic acid into the target, where
CC transposition at the predominant site indicates the presence of mismatch
CC at about that site. The methods are useful for typing a pathogenic
CC microorganism strain, for screening an embryo for the presence of
CC mutation, for detecting the presence of known mutation in a gene of
CC interest, detecting the presence of a previously unidentified mutation in
CC a gene of interest, and diagnosing the presence or absence of a tumour-
CC promoting mutation. The kit is useful for detecting the presence of a
CC mutation or polymorphism of interest in a nucleic acid molecule. The
CC present sequence represents the Mu end DNA fragment construction
CC oligonucleotide MM1141.
XX
SQ Sequence 51 BP; 6 A; 13 C; 11 G; 21 T; 0 U; 0 Other;

Query Match 78.8%; Score 39.4; DB 13; Length 51;
Best Local Similarity 97.6%; Pred. No. 7.5e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
| ||||||||||||||||||||||||||||||||||||||||||
Db 42 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 2

RESULT 6
AAD21279
ID AAD21279 standard; DNA; 54 BP.
XX
AC AAD21279;
XX
DT 11-SEP-2003 (revised)
DT 28-JAN-2002 (first entry)
XX
DE Precut transposon end of Bacteriophage Mu non-transferred strand.
XX
KW Insertional mutation; synaptic complex; transposon; screening; ds.
XX
OS Enterobacteria phage Mu.
XX
PN US6294385-B1.
XX
PD 25-SEP-2001.
XX
PF 10-AUG-2000; 2000US-00635969.
XX
PR 23-SEP-1998; 98US-00159363.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Goryshin IY, Reznikoff WS;
XX
DR WPI; 2001-656171/75.
XX

```
PT   Making an insertional mutations, especially useful for efficiently
PT   inserting a transposable polynucleotide in a target cell, comprises
PT   introducing into the target cell a synaptic complex.
XX
PS   Claim 6; Col 2; 11pp; English.
XX
CC   The present invention relates to a method for making an insertional
CC   mutation at a random or quasi-random position in cellular nucleic acid in
CC   a target cell comprising introducing into the target cell a synaptic
CC   complex. The method is particularly useful for efficiently inserting a
CC   transposable polynucleotide at random or quasi-random locations in the
CC   chromosomal or extra-chromosomal nucleic acid of a target cell. The
CC   method may also be used for screening the genome of cells that comprise
CC   an insertional mutation that induces a phenotypic or genotypic change
CC   relative to the cells that are not subject to insertional mutagenesis.
CC   The present sequence is the precut transposon end of Bacteriophage Mu non
CC   -transferred strand, used in the exemplification of the invention.
CC   (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ   Sequence 54 BP; 21 A; 12 C; 14 G; 7 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 4;  Length 54;
  Best Local Similarity   97.6%;  Pred. No. 7.5e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         14 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 7
ACF58171
ID   ACF58171 standard; DNA; 54 BP.
XX
AC   ACF58171;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Cat-Mu(NotI) transposon modified end fragment.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Synthetic.
OS   Bacteriophage mu.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .4
FT                   /*tag=  a
FT                   /note= "the 5' end of this strand overhangs the 3'end of
FT                   the complementary strand"
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Example 4; Fig 2; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a modified Cat-Mu transposon
CC   containing Mu ends with engineered NotI restriction site
XX
SQ   Sequence 54 BP; 19 A; 14 C; 14 G; 7 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 10;  Length 54;
  Best Local Similarity   97.6%;  Pred. No. 7.5e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         14 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 8
ACF58170
ID   ACF58170 standard; DNA; 54 BP.
XX
AC   ACF58170;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Cat-Mu transposon containing wild-type Mu ends.
XX
KW   Transposon; genetic engineering; transposase; Cat-Mu; ds.
XX
OS   Bacteriophage mu.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .4
FT                   /*tag=  a
FT                   /note= "the 5' end of this strand overhangs the 3'end of
FT                   the complementary strand"
XX
PN   WO2003087370-A1.
XX
PD   23-OCT-2003.
XX
PF   14-APR-2003; 2003WO-FI000285.
XX
PR   18-APR-2002; 2002FI-00000746.
XX
PA   (FINN-) FINNZYMES OY.
XX
PI   Savilahti H,  Tieaho V;
XX
DR   WPI; 2003-845329/78.
XX
PT   New transposon nucleic acid comprising a genetically engineered
PT   translation stop signal within a transposon end sequence recognized by a
PT   transposase useful for producing deletion derivatives of polypeptide.
XX
PS   Example 4; Fig 2; 0pp; English.
XX
CC   The invention relates to a transposon nucleic acid comprising a
CC   genetically engineered translation stop signal in the three reading
CC   frames at least partly within a transposon end sequence recognized by a
CC   transposase. The transposon is useful for producing deletion derivatives
CC   of polypeptide coding nucleic acids. The method involves performing a
CC   transposition reaction in the presence of a target nucleic acid
CC   containing a polypeptide coding nucleic acid and in the presence of a
CC   transposon containing a genetically engineered translation stop signal
CC   sequence in three reading frames at least partly within a transposon end
CC   sequence recognized by a transposase; and recovering a target nucleic
```

```
CC   acid having the transposon incorporated in the protein coding nucleic
CC   acid. The present sequence represents a Cat-Mu transposon containing wild
CC   -type Mu ends
XX
SQ   Sequence 54 BP; 21 A; 12 C; 14 G; 7 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 10;  Length 54;
  Best Local Similarity   97.6%;  Pred. No. 7.5e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         14 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 54


RESULT 9
ADS13820
ID   ADS13820 standard; DNA; 56 BP.
XX
AC   ADS13820;
XX
DT   16-DEC-2004  (first entry)
XX
DE   Mu end DNA fragment construction oligonucleotide MM1138.
XX
KW   ss; mismatch detection; transposition detection; pathogen typing;
KW   embryo screening; mutation detection; Mu end DNA; MM1138.
XX
OS   Enterobacteria phage Mu.
OS   Synthetic.
XX
PN   US2004191821-A1.
XX
PD   30-SEP-2004.
XX
PF   26-MAR-2004; 2004US-00809688.
XX
PR   28-MAR-2003; 2003US-0457934P.
XX
PA   (USSA ) US SEC OF ARMY.
XX
PI   Yanagihara K,  Mizuuchi K;
XX
DR   WPI; 2004-689846/67.
XX
PT   Detecting a mismatch in a test double stranded nucleic acid target,
PT   useful for typing a pathogenic microorganism strain, comprises detecting
PT   transposition of the Mu-end nucleic acid into the target.
XX
PS   Example 1; SEQ ID NO 1; 24pp; English.
XX
CC   The invention relates to a method of detecting a mismatch in a test
CC   double stranded nucleic acid target which comprises detecting
CC   transposition of the Mu-end nucleic acid into the target, where
CC   transposition at the predominant site indicates the presence of mismatch
CC   at about that site. The methods are useful for typing a pathogenic
CC   microorganism strain, for screening an embryo for the presence of
CC   mutation, for detecting the presence of known mutation in a gene of
CC   interest, detecting the presence of a previously unidentified mutation in
CC   a gene of interest, and diagnosing the presence or absence of a tumour-
CC   promoting mutation. The kit is useful for detecting the presence of a
CC   mutation or polymorphism of interest in a nucleic acid molecule. The
CC   present sequence represents the Mu end DNA fragment construction
CC   oligonucleotide MM1138.
XX
SQ   Sequence 56 BP; 22 A; 12 C; 15 G; 7 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 13;  Length 56;
  Best Local Similarity   97.6%;  Pred. No. 7.6e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         15 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 55


RESULT 10
AAZ28883/c
ID   AAZ28883 standard; cDNA; 58 BP.
XX
AC   AAZ28883;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Mini-Mu transposable element deletion region delta-1.
XX
KW   Transposon Mu; transposable element; fusion protein; attachment site;
KW   attL; attR; protein domain library; enzyme; accelerated evolution; ss.
XX
OS   Synthetic.
XX
PN   US5981177-A.
XX
PD   09-NOV-1999.
XX
PF   25-JAN-1995;   95US-00378548.
XX
PR   25-JAN-1995;   95US-00378548.
XX
PA   (DEMI/) DEMIRJIAN D C.
PA   (CASA/) CASADABAN M J.
PA   (WEBE/) WEBER J M.
PA   (GAIN/) GAINES G L.
XX
PI   Casadaban MJ,  Demirjian DC,  Weber JM,  Gaines GL;
XX
DR   WPI; 1999-633307/54.
XX
PT   Generating fusion proteins using transposable elements, useful for
PT   development of a protein domain library and in the construction of multi-
PT   functional enzymes.
XX
PS   Example 2; Fig 2; 41pp; English.
XX
CC   The invention relates to a Mu-like transposable element (I) used for
CC   generating functional fusion proteins after insertion into a target DNA.
CC   The Mu-like element comprises: (a) a left transposable element attachment
CC   site attL and a right transposable element attachment site attR, where
CC   attR is no more than 62 nucleotides long; (b) a site for insertion of an
CC   exogenous DNA sequence encoding for a protein domain located between attL
CC   and attR; (c) after insertion of the transposable element into a target
CC   DNA sequence, a fusion mRNA sequence is transcribed originating either
CC   from the target DNA on either side of the transposable element or from
CC   inside the transposable element and continuing through the attachment
CC   site sequences and into the protein coding region, resulting in a single
CC   fusion open reading frame (ORF). The constructs are useful in the
CC   development of a protein domain library, in the construction of multi-
CC   functional enzymes and in the accelerated evolution of new enzymatic
CC   activities. The sequences AAZ28880-Z28884 represent deletion mini-Mu
CC   elements of the invention (encoded ORF - AAY55901-Y55906)
XX
SQ   Sequence 58 BP; 8 A; 15 C; 13 G; 22 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 2;  Length 58;
  Best Local Similarity   97.6%;  Pred. No. 7.6e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         49 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 9


RESULT 11
AAZ28882/c
ID   AAZ28882 standard; cDNA; 71 BP.
XX
```

```
AC   AAZ28882;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Mini-Mu transposable element deletion region delta-21.
XX
KW   Transposon Mu; transposable element; fusion protein; attachment site;
KW   attL; attR; protein domain library; enzyme; accelerated evolution; ss.
XX
OS   Synthetic.
XX
PN   US5981177-A.
XX
PD   09-NOV-1999.
XX
PF   25-JAN-1995;   95US-00378548.
XX
PR   25-JAN-1995;   95US-00378548.
XX
PA   (DEMI/) DEMIRJIAN D C.
PA   (CASA/) CASADABAN M J.
PA   (WEBE/) WEBER J M.
PA   (GAIN/) GAINES G L.
XX
PI   Casadaban MJ,  Demirjian DC,  Weber JM,  Gaines GL;
XX
DR   WPI; 1999-633307/54.
XX
PT   Generating fusion proteins using transposable elements, useful for
PT   development of a protein domain library and in the construction of multi-
PT   functional enzymes.
XX
PS   Example 2; Fig 2; 41pp; English.
XX
CC   The invention relates to a Mu-like transposable element (I) used for
CC   generating functional fusion proteins after insertion into a target DNA.
CC   The Mu-like element comprises: (a) a left transposable element attachment
CC   site attL and a right transposable element attachment site attR, where
CC   attR is no more than 62 nucleotides long; (b) a site for insertion of an
CC   exogenous DNA sequence encoding for a protein domain located between attL
CC   and attR; (c) after insertion of the transposable element into a target
CC   DNA sequence, a fusion mRNA sequence is transcribed originating either
CC   from the target DNA on either side of the transposable element or from
CC   inside the transposable element and continuing through the attachment
CC   site sequences and into the protein coding region, resulting in a single
CC   fusion open reading frame (ORF). The constructs are useful in the
CC   development of a protein domain library, in the construction of multi-
CC   functional enzymes and in the accelerated evolution of new enzymatic
CC   activities. The sequences AAZ28880-Z28884 represent deletion mini-Mu
CC   elements of the invention (encoded ORF - AAY55901-Y55906)
XX
SQ   Sequence 71 BP; 14 A; 18 C; 16 G; 23 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 2;  Length 71;
  Best Local Similarity   97.6%;  Pred. No. 7.8e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         62 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 22


RESULT 12
ACC80745
ID   ACC80745 standard; DNA; 82 BP.
XX
AC   ACC80745;
XX
DT   15-OCT-2003  (first entry)
XX
DE   Transposon-based targeting construct related primer Mu2-Neo-2.
XX
KW   Targeting construct; targeting vector; transposon; recombination;
KW   deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS   Synthetic.
XX
PN   WO2003031629-A1.
XX
PD   17-APR-2003.
XX
PF   08-OCT-2002; 2002WO-AU001367.
XX
PR   09-OCT-2001; 2001AU-00008174.
PR   23-MAY-2002; 2002AU-00002522.
XX
PA   (COPY-) COPYRAT PTY LTD.
XX
PI   Morrison J,  Zhang C;
XX
DR   WPI; 2003-393445/37.
XX
PT   Preparing a targeting construct using a transposon and DNA recombination
PT   sequence, useful in making a targeting vector capable of modifying plant
PT   and/or animal genome in a predetermined way.
XX
PS   Example 5; Page 51; 92pp; English.
XX
CC   The invention relates to methods of preparing a targeting construct for
CC   use in a targeting vector capable of modifying a target DNA sequence, by
CC   obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC   sequence comprising a transposon sequence and a DNA recombination
CC   sequence at two sites in the copy of the target DNA sequence, and
CC   inducing a recombination event between the recombination sequences to
CC   delete a portion of the copy of the target DNA sequence. The methods and
CC   compositions of the present invention are useful for preparing a target
CC   construct for use in a targeting vector for gene targeting or homologous
CC   recombination. They can also be used for precisely modifying plant and/or
CC   animal genome in a predetermined way. This sequence represents a primer
CC   used in an example of the invention
XX
SQ   Sequence 82 BP; 35 A; 21 C; 16 G; 10 T; 0 U; 0 Other;

  Query Match             78.8%;  Score 39.4;  DB 10;  Length 82;
  Best Local Similarity   97.6%;  Pred. No. 8e-05;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | ||||||||||||||||||||||||||||||||||||||||
Db         21 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 61


RESULT 13
AAZ28881/c
ID   AAZ28881 standard; cDNA; 83 BP.
XX
AC   AAZ28881;
XX
DT   15-FEB-2000  (first entry)
XX
DE   Mini-Mu transposable element deletion region delta-66.
XX
KW   Transposon Mu; transposable element; fusion protein; attachment site;
KW   attL; attR; protein domain library; enzyme; accelerated evolution; ss.
XX
OS   Synthetic.
XX
PN   US5981177-A.
XX
PD   09-NOV-1999.
XX
PF   25-JAN-1995;   95US-00378548.
XX
PR   25-JAN-1995;   95US-00378548.
XX
PA   (DEMI/) DEMIRJIAN D C.
```

PA (CASA/) CASADABAN M J.
PA (WEBE/) WEBER J M.
PA (GAIN/) GAINES G L.
XX
PI Casadaban MJ, Demirjian DC, Weber JM, Gaines GL;
XX
DR WPI; 1999-633307/54.
XX
PT Generating fusion proteins using transposable elements, useful for
PT development of a protein domain library and in the construction of multi-
PT functional enzymes.
XX
PS Example 2; Fig 2; 41pp; English.
XX
CC The invention relates to a Mu-like transposable element (I) used for
CC generating functional fusion proteins after insertion into a target DNA.
CC The Mu-like element comprises: (a) a left transposable element attachment
CC site attL and a right transposable element attachment site attR, where
CC attR is no more than 62 nucleotides long; (b) a site for insertion of an
CC exogenous DNA sequence encoding for a protein domain located between attL
CC and attR; (c) after insertion of the transposable element into a target
CC DNA sequence, a fusion mRNA sequence is transcribed originating either
CC from the target DNA on either side of the transposable element or from
CC inside the transposable element and continuing through the attachment
CC site sequences and into the protein coding region, resulting in a single
CC fusion open reading frame (ORF). The constructs are useful in the
CC development of a protein domain library, in the construction of multi-
CC functional enzymes and in the accelerated evolution of new enzymatic
CC activities. The sequences AAZ28880-Z28884 represent deletion mini-Mu
CC elements of the invention (encoded ORF - AAY55901-Y55906)
XX
SQ Sequence 83 BP; 17 A; 19 C; 18 G; 29 T; 0 U; 0 Other;

Query Match 78.8%; Score 39.4; DB 2; Length 83;
Best Local Similarity 97.6%; Pred. No. 8e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
| ||||||||||||||||||||||||||||||||||||||||||
Db 74 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 34

RESULT 14
ACC80741
ID ACC80741 standard; DNA; 83 BP.
XX
AC ACC80741;
XX
DT 15-OCT-2003 (first entry)
XX
DE Transposon-based targeting construct related primer Mu1-4.
XX
KW Targeting construct; targeting vector; transposon; recombination;
KW deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS Synthetic.
XX
PN WO2003031629-A1.
XX
PD 17-APR-2003.
XX
PF 08-OCT-2002; 2002WO-AU001367.
XX
PR 09-OCT-2001; 2001AU-00008174.
PR 23-MAY-2002; 2002AU-00002522.
XX
PA (COPY-) COPYRAT PTY LTD.
XX
PI Morrison J, Zhang C;
XX
DR WPI; 2003-393445/37.
XX
PT Preparing a targeting construct using a transposon and DNA recombination
PT sequence, useful in making a targeting vector capable of modifying plant
PT and/or animal genome in a predetermined way.
XX
PS Example 5; Page 50; 92pp; English.
XX
CC The invention relates to methods of preparing a targeting construct for
CC use in a targeting vector capable of modifying a target DNA sequence, by
CC obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC sequence comprising a transposon sequence and a DNA recombination
CC sequence at two sites in the copy of the target DNA sequence, and
CC inducing a recombination event between the recombination sequences to
CC delete a portion of the copy of the target DNA sequence. The methods and
CC compositions of the present invention are useful for preparing a target
CC construct for use in a targeting vector for gene targeting or homologous
CC recombination. They can also be used for precisely modifying plant and/or
CC animal genome in a predetermined way. This sequence represents a primer
CC used in an example of the invention
XX
SQ Sequence 83 BP; 29 A; 20 C; 18 G; 16 T; 0 U; 0 Other;

Query Match 78.8%; Score 39.4; DB 10; Length 83;
Best Local Similarity 97.6%; Pred. No. 8e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
| ||||||||||||||||||||||||||||||||||||||||||
Db 24 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 64

RESULT 15
ACC80738
ID ACC80738 standard; DNA; 84 BP.
XX
AC ACC80738;
XX
DT 15-OCT-2003 (first entry)
XX
DE Transposon-based targeting construct related primer Mu1-1.
XX
KW Targeting construct; targeting vector; transposon; recombination;
KW deletion; plant genome; animal genome; primer; PCR; ss.
XX
OS Synthetic.
XX
PN WO2003031629-A1.
XX
PD 17-APR-2003.
XX
PF 08-OCT-2002; 2002WO-AU001367.
XX
PR 09-OCT-2001; 2001AU-00008174.
PR 23-MAY-2002; 2002AU-00002522.
XX
PA (COPY-) COPYRAT PTY LTD.
XX
PI Morrison J, Zhang C;
XX
DR WPI; 2003-393445/37.
XX
PT Preparing a targeting construct using a transposon and DNA recombination
PT sequence, useful in making a targeting vector capable of modifying plant
PT and/or animal genome in a predetermined way.
XX
PS Example 5; Page 50; 92pp; English.
XX
CC The invention relates to methods of preparing a targeting construct for
CC use in a targeting vector capable of modifying a target DNA sequence, by
CC obtaining a copy of the target DNA sequence in vitro, inserting a DNA
CC sequence comprising a transposon sequence and a DNA recombination
CC sequence at two sites in the copy of the target DNA sequence, and
CC inducing a recombination event between the recombination sequences to
CC delete a portion of the copy of the target DNA sequence. The methods and
CC compositions of the present invention are useful for preparing a target

CC construct for use in a targeting vector for gene targeting or homologous
CC recombination. They can also be used for precisely modifying plant and/or
CC animal genome in a predetermined way. This sequence represents a primer
CC used in an example of the invention
XX
SQ Sequence 84 BP; 29 A; 20 C; 19 G; 16 T; 0 U; 0 Other;

Query Match 78.8%; Score 39.4; DB 10; Length 84;
Best Local Similarity 97.6%; Pred. No. 8e-05;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
| ||||||||||||||||||||||||||||||||||||||||||
Db 24 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 64

Search completed: January 17, 2006, 19:35:11
Job time : 32.092 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 211.625 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-5
Perfect score: 50
Sequence: 1 tgattgattgaacgaaaaac.....ttcacgataaatgcgaaaac 50

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	40.4	80.8	370	10	CZ569838	CZ569838 OB_Ba000
	2	39.4	78.8	226	9	CC742071	CC742071 ZMMBBb011
	3	39.4	78.8	589	10	CL858345	CL858345 OR_CBa008
	4	36.2	72.4	543	10	CZ821413	CZ821413 OC_Ba019
	5	36.2	72.4	688	10	CZ820900	CZ820900 OC_Ba019
	6	29.2	58.4	247	10	CZ821171	CZ821171 OC_Ba019
	7	26.6	53.2	582	3	BJ016030	BJ016030 BJ016030
c	8	26.6	53.2	802	10	DU032104	DU032104 13515 Tom
c	9	26.2	52.4	215	3	BI953982	BI953982 HVSMEm001
c	10	26.2	52.4	771	3	BI953934	BI953934 HVSMEm001
c	11	25.8	51.6	757	3	BI953957	BI953957 HVSMEm001
c	12	25.6	51.2	196	3	BI953931	BI953931 HVSMEm001
c	13	25.6	51.2	752	3	BI953878	BI953878 HVSMEm001
c	14	25.6	51.2	771	3	BI953902	BI953902 HVSMEm001
	15	25.6	51.2	973	10	DU005263	DU005263 300893 To
c	16	25.4	50.8	484	9	AZ152875	AZ152875 SP_0046_B
c	17	25.4	50.8	887	7	CN586362	CN586362 USDA-FP_1
c	18	25.2	50.4	274	8	T31184	T31184 EST28439 Hu
	19	25.2	50.4	297	6	CF504687	CF504687 USDA-FP_1
	20	25.2	50.4	491	8	DN797614	DN797614 USDA-FP_A
c	21	25.2	50.4	737	3	BI954000	BI954000 HVSMEm001
c	22	25.2	50.4	741	3	BI953998	BI953998 HVSMEm001
c	23	25.2	50.4	748	3	BI954023	BI954023 HVSMEm001
	24	25.2	50.4	776	6	CB292852	CB292852 UCRCS01_0
	25	25	50.0	346	3	BI507751	BI507751 BB170007B
	26	24.8	49.6	705	3	BQ149276	BQ149276 NF088G08F
c	27	24.6	49.2	248	8	F05165	F05165 HSC02H021 n
c	28	24.6	49.2	304	1	AA481614	AA481614 aa35g09.r
c	29	24.6	49.2	318	5	C14524	C14524 C14524 Clon
c	30	24.6	49.2	403	1	AA262843	AA262843 zs24d05.r
c	31	24.6	49.2	535	5	C14413	C14413 C14413 Clon
c	32	24.6	49.2	561	1	AW954984	AW954984 EST367054
c	33	24.6	49.2	590	9	AQ998995	AQ998995 RPCI-23-2
c	34	24.6	49.2	634	1	AW955329	AW955329 EST367399
c	35	24.6	49.2	731	1	AL041260	AL041260 DKFZp434L
	36	24.6	49.2	753	6	CF449523	CF449523 EST685868
c	37	24.6	49.2	766	3	BI953900	BI953900 HVSMEm001
c	38	24.6	49.2	856	3	BI772228	BI772228 603056184
c	39	24.6	49.2	874	8	DR952125	DR952125 EST114366
c	40	24.6	49.2	994	3	BM804853	BM804853 AGENCOURT
	41	24.4	48.8	592	9	AZ176111	AZ176111 SP_0140_B
c	42	24.4	48.8	671	9	BZ833089	BZ833089 CH240_222
c	43	24.4	48.8	703	5	BW085322	BW085322 BW085322
	44	24.4	48.8	706	3	BM617600	BM617600 170006871
	45	24.4	48.8	825	11	CR792741	CR792741 GR0AAA1BE

ALIGNMENTS

RESULT 1
CZ569838
LOCUS CZ569838 370 bp DNA linear GSS 20-JUN-2005
DEFINITION OB__Ba0003F02.r OB__Ba Oryza brachyantha genomic clone
OB__Ba0003F02 3', genomic survey sequence.
ACCESSION CZ569838
VERSION CZ569838.1 GI:68013639
KEYWORDS GSS.
SOURCE Oryza brachyantha
ORGANISM Oryza brachyantha
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 370)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Insert Length: 145 Std Error: 0.00
Plate: 0003 row: F column: 02
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .370
/organism="Oryza brachyantha"
/mol_type="genomic DNA"
/db_xref="taxon:4533"
/clone="OB__Ba0003F02"
/tissue_type="leaves"
/dev_stage="mature"
/lab_host="DH10B"
/clone_lib="OB__Ba"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

```
  Query Match             80.8%;  Score 40.4;  DB 10;  Length 370;
  Best Local Similarity   97.6%;  Pred. No. 0.00023;
  Matches   41;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          9 TGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              || |||||||||||||||||||||||||||||||||||||||
Db         85 TGCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 126


RESULT 2
CC742071
LOCUS       CC742071                 226 bp    DNA     linear   GSS 25-JUN-2003
DEFINITION  ZMMBBb0115J07.f ZMMBBb Zea mays genomic clone ZMMBBb0115J07 5',
            genomic survey sequence.
ACCESSION   CC742071
VERSION     CC742071.1  GI:32194524
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 226)
  AUTHORS   Yu,Y., Kim,H.R., Hatfield,J., Soderlund,C., Bharti,A.K., Messing,J.
            and Wing,R.
  TITLE     Sequencing of the maize genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: M13r
            Plate: 0115  row: J  column: 07
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .226
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBBb0115J07"
                     /lab_host="DH10B"
                     /clone_lib="ZMMBBb"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Zea mays L. ssp. mays"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 9;  Length 226;
  Best Local Similarity   97.6%;  Pred. No. 0.00053;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | |||||||||||||||||||||||||||||||||||||||
Db        182 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 222


RESULT 3
CL858345
LOCUS       CL858345                 589 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089D12.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089D12
            5', genomic survey sequence.
ACCESSION   CL858345
VERSION     CL858345.1  GI:51269584
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 589)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: D  column: 12
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .589
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089D12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             78.8%;  Score 39.4;  DB 10;  Length 589;
  Best Local Similarity   97.6%;  Pred. No. 0.00054;
  Matches   40;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | |||||||||||||||||||||||||||||||||||||||
Db        495 GCACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 535


RESULT 4
CZ821413
LOCUS       CZ821413                 543 bp    DNA     linear   GSS 26-JUL-2005
DEFINITION  OC__Ba0199P13.r OC__Ba Oryza coarctata genomic clone OC__Ba0199P13
            3', genomic survey sequence.
ACCESSION   CZ821413
VERSION     CZ821413.1  GI:71261266
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 543)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
```

```
           BACKWARD: CAC TCA TTA GGC ACC CCA
           Plate: 0199  row: P  column: 13
           Seq primer: CAC TCA TTA GGC ACC CCA
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .543
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0199P13"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match           72.4%;  Score 36.2;  DB 10;  Length 543;
  Best Local Similarity   92.7%;  Pred. No. 0.0083;
  Matches   38;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | |||||||| |||||||||||||||||||| ||||||||||
Db        473 GCACGAAAAATGCGAAAGCGTTTCACGATACATGCGAAAAC 513


RESULT 5
CZ820900
LOCUS       CZ820900                688 bp    DNA     linear   GSS 26-JUL-2005
DEFINITION  OC__Ba0199D21.r OC__Ba Oryza coarctata genomic clone OC__Ba0199D21
            3', genomic survey sequence.
ACCESSION   CZ820900
VERSION     CZ820900.1  GI:71260753
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0199  row: D  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0199D21"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match           72.4%;  Score 36.2;  DB 10;  Length 688;
  Best Local Similarity   92.7%;  Pred. No. 0.0083;
  Matches   38;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | |||||||| |||||||||||||||||||| ||||||||||
Db        440 GCACGAAAAACGCGACAGCGTCTCACGATAAATGCGAAAAC 480


RESULT 6
CZ821171
LOCUS       CZ821171                247 bp    DNA     linear   GSS 26-JUL-2005
DEFINITION  OC__Ba0199J21.r OC__Ba Oryza coarctata genomic clone OC__Ba0199J21
            3', genomic survey sequence.
ACCESSION   CZ821171
VERSION     CZ821171.1  GI:71261024
KEYWORDS    GSS.
SOURCE      Oryza coarctata (Porteresia coarctata)
  ORGANISM  Oryza coarctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Kim,H., Collura,K., Wissotski,M., Byrne,M., Stum,D., Smart,D.,
            Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Soderlund,C. and
            Wing,R.
  TITLE     OMAP (Oryza Map Alignment Project)- Arizona Genomics Institute
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: rwing@genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0199  row: J  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Oryza coarctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:77588"
                     /clone="OC__Ba0199J21"
                     /tissue_type="leaves"
                     /dev_stage="mature"
                     /lab_host="DH10B"
                     /clone_lib="OC__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match           58.4%;  Score 29.2;  DB 10;  Length 247;
  Best Local Similarity   81.0%;  Pred. No. 3.2;
  Matches   34;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          9 TGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              || |  |||||  |||||||||||||| ||| |||| ||||||
Db        143 TGCATCAAAAATCCGAAAGCGTTTCATGATCAATGTGAAAAC 184


RESULT 7
BJ016030
LOCUS       BJ016030                582 bp    mRNA    linear   EST 05-DEC-2001
DEFINITION  BJ016030 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA008A02 3',
            mRNA sequence.
ACCESSION   BJ016030
VERSION     BJ016030.1  GI:17375178
KEYWORDS    EST.
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
```

```
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
  TITLE     Medaka EST Project in Takeda's lab
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Oryzias latipes"
                     /mol_type="mRNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="MF01SSA008A02"
                     /sex="mixture of female and male"
                     /tissue_type="whole embryo"
                     /dev_stage="segmentation stage 20 - 25"
                     /clone_lib="MF01SSA cDNA"
ORIGIN

  Query Match             53.2%;  Score 26.6;  DB 3;  Length 582;
  Best Local Similarity   71.4%;  Pred. No. 30;
  Matches   35;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          1 TGATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAA 49
              | ||| ||| ||||  |||  | ||||| |||||  |||||| | || |
Db        210 TCATTTATTAAACGTGAAAACCAAAAGCATTTCAGAATAAATACTAACA 258


RESULT 8
DU032104/c
LOCUS       DU032104                802 bp    DNA     linear   GSS 12-AUG-2005
DEFINITION  13515 Tomato HindIII BAC Library Lycopersicon esculentum genomic
            clone LE_HBa0169A20 3, genomic survey sequence.
ACCESSION   DU032104
VERSION     DU032104.1  GI:72445405
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 802)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 13518
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 169  row: A  column: 20
            Seq primer: SP6
            Class: BAC ends
            High quality sequence start: 36
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .802
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="LE_HBa0169A20"
                     /lab_host="E. coli"
                     /clone_lib="Tomato HindIII BAC Library"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             53.2%;  Score 26.6;  DB 10;  Length 802;
  Best Local Similarity   71.4%;  Pred. No. 30;
  Matches   35;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 GATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              |||  ||| || ||||||   | |||  |||||| | | ||| ||||||
Db        693 GATGAATTAAAAGAAAAAGAAGCAAGATTTTCACTAAATATGAGAAAAC 645


RESULT 9
BI953982/c
LOCUS       BI953982                215 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0015K12f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
            clone HVSMEm0015K12f, mRNA sequence.
ACCESSION   BI953982
VERSION     BI953982.1  GI:16299103
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 215)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
            Simmons,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 174
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 4
            High quality sequence stop: 216.
FEATURES             Location/Qualifiers
     source          1. .215
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEm0015K12f"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
```

Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 52.4%; Score 26.2; DB 3; Length 215;
Best Local Similarity 79.5%; Pred. No. 41;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAA 48
              | ||||||||||||||||||| ||||  |||    |||
Db         40 GCACGAAAAACGCGAAAGCGTTGCACGTAAAAGCGAAAA 2
```

RESULT 10
BI953934/c
LOCUS BI953934 771 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEm0015G22f Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA clone HVSMEm0015G22f, mRNA sequence.
ACCESSION BI953934
VERSION BI953934.1 GI:16299017
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 771)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 471
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 25
High quality sequence stop: 736.

FEATURES Location/Qualifiers
source 1. .771
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEm0015G22f"
/tissue_type="green seedling leaf"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Morex (mla) plants were greenhouse grown in the R Wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 52.4%; Score 26.2; DB 3; Length 771;
Best Local Similarity 79.5%; Pred. No. 42;
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAA 48
              | ||||||||||||||||| |||||||  |||    |||
Db         40 GCACGAAAAACGCGAAAGCTTTTCACGTAAAAGCAAAAA 2
```

RESULT 11
BI953957/c
LOCUS BI953957 757 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEm0015I12f Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA clone HVSMEm0015I12f, mRNA sequence.
ACCESSION BI953957
VERSION BI953957.1 GI:16299057
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 757)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 590
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 4
High quality sequence stop: 756.

FEATURES Location/Qualifiers
source 1. .757
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"

/db_xref="taxon:112509"
/clone="HVSMEm0015I12f"
/tissue_type="green seedling leaf"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Morex (mla) plants were greenhouse grown in the R Wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 51.6%; Score 25.8; DB 3; Length 757;
Best Local Similarity 92.7%; Pred. No. 60;
Matches 38; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAAC 50
              | |||||||||||||||| ||||||| |||||||||||||
Db         40 GCACGAAAAACGCGAAAGCTTTTCACG-TAAATGCGAAAAC 1
```

RESULT 12
BI953931/c
LOCUS BI953931 196 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEm0015G18f Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA clone HVSMEm0015G18f, mRNA sequence.
ACCESSION BI953931
VERSION BI953931.1 GI:16299011
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 196)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Blumeria infected Morex (compatible) seedling cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu

Total hq bases = 164
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 6
High quality sequence stop: 197.

FEATURES Location/Qualifiers
source 1. .196
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEm0015G18f"
/tissue_type="green seedling leaf"
/lab_host="TJC121"
/clone_lib="Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected)"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Morex (mla) plants were greenhouse grown in the R Wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

ORIGIN

Query Match 51.2%; Score 25.6; DB 3; Length 196;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 28; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAA 41
              | ||||||||||||||||||||||||||  | |
Db         32 GCACGAAAAACGCGAAAGCGTTTCACGTAATA 1
```

RESULT 13
BI953878/c
LOCUS BI953878 752 bp mRNA linear EST 19-OCT-2001
DEFINITION HVSMEm0015C18f Hordeum vulgare green seedling EST library HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA clone HVSMEm0015C18f, mRNA sequence.
ACCESSION BI953878
VERSION BI953878.1 GI:16298915
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 752)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D., Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource

```
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 537
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 15
            High quality sequence stop: 729.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEm0015C18f"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN

  Query Match             51.2%;  Score 25.6;  DB 3;  Length 752;
  Best Local Similarity   87.5%;  Pred. No. 71;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAA 41
              | |||||||||||||||| |||||||  |||
Db         32 GCACGAAAAACGCGAAAGCTTTTCACGTAAAA 1


RESULT 14
BI953902/c
LOCUS       BI953902                771 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0015E24f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare subsp. vulgare cDNA
            clone HVSMEm0015E24f, mRNA sequence.
ACCESSION   BI953902
VERSION     BI953902.1  GI:16298957
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
            Simmons,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 486
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 16
            High quality sequence stop: 619.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="Morex"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HVSMEm0015E24f"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
ORIGIN

  Query Match             51.2%;  Score 25.6;  DB 3;  Length 771;
  Best Local Similarity   87.5%;  Pred. No. 71;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         10 GAACGAAAAACGCGAAAGCGTTTCACGATAAA 41
              | |||||||||||||||| |||||||  |||
Db         32 GCACGAAAAACGCGAAAGCTTTTCACGTAAAA 1
```

```
RESULT 15
DU005263
LOCUS       DU005263                 973 bp    DNA     linear   GSS 12-AUG-2005
DEFINITION  300893 Tomato MboI BAC Library Lycopersicon esculentum genomic
            clone SL_MboI0044G12 3, genomic survey sequence.
ACCESSION   DU005263
VERSION     DU005263.1  GI:72402335
KEYWORDS    GSS.
SOURCE      Lycopersicon esculentum (Solanum lycopersicum)
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Mueller,L.A., Buels,R.M., Wang,Y., Tanksley,S.D., Giovannoni,J.J.,
            Van Eck,J. and Stack,S.
  TITLE     BAC end sequencing from three Solanum lycopersicon libraries
  JOURNAL   Unpublished (2005)
COMMENT     Other_GSSs: 240458
            Contact: Lukas Mueller
            Tanksley Lab, Dept. of Plant Breeding
            Cornell University
            251 Emerson Hall, Ithaca, NY 14853, USA
            Tel: 607-255-6557
            Fax: 607-255-6683
            Email: sgn-feedback@sgn.cornell.edu
            Plate: 44  row: G  column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence start: 50
            High quality sequence stop: 416.
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Lycopersicon esculentum"
                     /mol_type="genomic DNA"
                     /cultivar="Heinz 1706"
                     /db_xref="taxon:4081"
                     /clone="SL_MboI0044G12"
                     /lab_host="E. coli"
                     /clone_lib="Tomato MboI BAC Library"
                     /note="Vector: pBeloBAC11; Site_1: MboI"
ORIGIN

  Query Match             51.2%;  Score 25.6;  DB 10;  Length 973;
  Best Local Similarity   70.8%;  Pred. No. 71;
  Matches   34;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          2 GATTGATTGAACGAAAAACGCGAAAGCGTTTCACGATAAATGCGAAAA 49
              ||||||||||   | ||| |||   |  ||   ||||||||| ||||
Db        454 GATTGATTGAATTGACAACTCGACCTCACTTATAGATAAATGCTAAAA 501


Search completed: January 18, 2006, 11:40:50
Job time : 212.625 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:06:53 ; Search time 24804 Seconds
(without alignments)
11032.288 Million cell updates/sec

Title: US-10-511-327-7
Perfect score: 4814
Sequence: 1 ggtaccctgtgaatgcgcaa.....taagatgcttttctgtgact 4814

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1636.8	34.0	15129	11	AY744149	AY744149 Dengue vi
	2	1636.8	34.0	15129	11	AY744150	AY744150 Dengue vi
	3	1636.8	34.0	15145	11	AY656170	AY656170 Dengue vi
	4	1636.8	34.0	15159	11	AY744148	AY744148 Dengue vi
	5	1636.8	34.0	15176	11	AY656169	AY656169 Dengue vi
	6	1636.8	34.0	15237	11	AY243469	AY243469 Chimeric
	7	1636.8	34.0	15239	11	AY243467	AY243467 Chimeric
	8	1636.8	34.0	15239	11	AY376438	AY376438 Dengue vi
	9	1636.8	34.0	15256	11	AY656168	AY656168 Chimeric
	10	1636.8	34.0	15268	11	AY243468	AY243468 Chimeric
	11	1636.8	34.0	15270	11	AY243466	AY243466 Chimeric
	12	1636.8	34.0	15270	11	AY648301	AY648301 Dengue vi
	13	1636.8	34.0	15287	11	AY656167	AY656167 Chimeric
c	14	1635.8	34.0	13561	11	AY028776	AY028776 TnphoZ mu
	15	1624	33.7	6343	11	SYNCLCTAPS	M88538 Expression
	16	1624	33.7	6571	11	SYNCLCTAPA	M88539 Expression
	17	1624	33.7	8071	11	SYNCOLCTAP	M88537 Expression
	18	1560.6	32.4	7311	6	CS091381	CS091381 Sequence
c	19	1469	30.5	5228	11	XXU25059	U25059 Cloning vec
c	20	1452	30.2	56167	3	AJ639924	AJ639924 Plasmid p
c	21	1450.4	30.1	11823	1	AY043299	AY043299 Aeromonas
	22	1450.4	30.1	11943	1	SYNPKF339	D45834 Size marker
	23	1448.2	30.1	3779	11	SYNPWT571V	L09155 pWT571 expr
	24	1447.8	30.1	4840	6	AX084356	AX084356 Sequence
	25	1447.8	30.1	5817	11	YRP7	U03501 Yeast repli
	26	1447.8	30.1	10667	11	YEP13	U03498 Yeast episo
	27	1447.8	30.1	10667	11	YEP213	U03499 Yeast episo
	28	1447	30.1	6303	11	AY093430	AY093430 Allelic e
	29	1447	30.1	6760	11	AY093429	AY093429 Allelic e
c	30	1446.8	30.1	3423	6	AR287564	AR287564 Sequence
c	31	1446.8	30.1	3474	6	A28084	A28084 pmTNF-MPH p
	32	1446.8	30.1	3474	6	A28085	A28085 pmTNF-MPH p
c	33	1446.8	30.1	3474	6	A75741	A75741 Sequence 10
c	34	1446.8	30.1	3474	6	AR085100	AR085100 Sequence
c	35	1446.8	30.1	3474	6	AR209770	AR209770 Sequence
c	36	1446.8	30.1	3474	6	AR287565	AR287565 Sequence
c	37	1446.8	30.1	3801	6	AR493833	AR493833 Sequence
	38	1446.8	30.1	4245	6	AR069365	AR069365 Sequence
	39	1446.8	30.1	4245	11	PACYC184	X06403 Cloning vec
c	40	1446.8	30.1	4633	6	A20355	A20355 plasmid pIG
	41	1446.8	30.1	4633	6	A20356	A20356 plasmid pIG
	42	1446.8	30.1	4753	11	AF129432	AF129432 Cloning v
c	43	1446.8	30.1	5201	6	AR493834	AR493834 Sequence
	44	1446.8	30.1	5281	11	SCU22104	U22104 Cloning vec
	45	1446.8	30.1	5443	11	AY796342	AY796342 Cloning v

ALIGNMENTS

RESULT 1
AY744149
LOCUS AY744149 15129 bp DNA circular SYN 03-DEC-2004
DEFINITION Dengue virus type 2 vector p2(delta30), complete sequence.
ACCESSION AY744149
VERSION AY744149.1 GI:56089726
KEYWORDS .
SOURCE Dengue virus type 2 vector p2(delta30)
ORGANISM Dengue virus type 2 vector p2(delta30)
other sequences; artificial sequences; vectors.
REFERENCE 1 (bases 1 to 15129)
AUTHORS Blaney,J.E. Jr., Hanson,C.T., Hanley,K.A., Murphy,B.R. and Whitehead,S.S.
TITLE Vaccine candidates derived from a novel infectious cDNA clone of an American genotype dengue virus type 2
JOURNAL (er) BMC Infect. Dis. 4 (1), 39 (2004)
PUBMED 15461822
REFERENCE 2 (bases 1 to 15129)
AUTHORS Whitehead,S.S.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2004) Laboratory of Infectious Diseases, NIAID, 50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES Location/Qualifiers
source 1. .15129
/organism="Dengue virus type 2 vector p2(delta30)"
/mol_type="other DNA"
/strain="Tonga/74"
/db_xref="taxon:301536"
/country="Tonga"
/focus
source 1. .10683
/organism="Dengue virus type 2"
/mol_type="other DNA"
/db_xref="taxon:11060"
source 10693. .15054
/organism="Cloning vector pBR322"
/mol_type="other DNA"
/db_xref="taxon:47470"
CDS 97. .10272
/codon_start=1
/product="polyprotein"

```
                     /protein_id="AAV70831.1"
                     /db_xref="GI:56089727"
                     /translation="MNNQRKKARNTPFNMLKRERNRVSTVQQLTKRFSLGMLQGRGPL
                     KLFMALVAFLRFLTIPPTAGILKRWGTIKKSKAINVLRGFRKEIGRMLNILNRRRRTV
                     GMIIMLTPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTKDGTNMCTLMAMDLGELCE
                     DTITYKCPFLKQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLE
                     TRTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRVLIFILLTAIA
                     PSMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPA
                     TLRKYCIEAKLTNTTTDSRCPTQGEPTLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGG
                     IVTCAMFTCKKNMEGKIVQPENLEYTVVITPHSGEEHAVGNDTGKHGKEVKITPQSSI
                     TEAELTGYGTVTMECSPRTGLDFNEMVLLQMEDKAWLVHRQWFLDLPLPWLPGADTQG
                     SNWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCR
                     LRMDKLQLKGMSYSMCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEK
                     RHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLDWFKKGSSIGQMF
                     ETTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIG
                     VIITWIGMNSRSTSLSVSLVLVGIVTLYLGVMVQADSGCVVSWKNKELKCGSGIFVTD
                     NVHTWTEQYKFQPESPSKLASAIQKAHEEGICGIRSVTRLENLMWKQITSELNHILSE
                     NEVKLTIMTGDIKGIMQVGKRSLRPQPTELRYSWKTWGKAKMLSTELHNQTFLIDGPE
                     TAECPNTNRAWNSLEVEDYGFGVFTTNIWLRLREKQDVFCDSKLMSAAIKDNRAVHAD
                     MGYWIESALNDTWKIEKASFIEVKSCHWPKSHTLWSNGVLESEMVIPKNFAGPVSQHN
                     NRPGYYTQTAGPWHLGKLEMDFDFCEGTTVVVTENCGNRGPSLRTTTASGKLITEWCC
                     RSCTLPPLRYRGEDGCWYGMEIRPLKEKEENLVSSLVTAGHGQIDNFSLGILGMALFL
                     EEMLRTRVGTKHAILLVAVSFVTLITGNMSFRDLGRVMVMVGATMTDDIGMGVTYLAL
                     LAAFRVRPTFAAGLLLRKLTSKELMMTTIGIVLLSQSSIPETILELTDALALGMMVLK
                     MVRNMEKYQLAVTIMAILCVPNAVILQNAWKVSCTILAVVSVSPLLLTSSQQKADWIP
                     LALTIKGLNPTAIFLTTLSRTNKKRSWPLNEAIMAVGMVSILASSLLKNDIPMTGPLV
                     AGGLLTVCYVLTGRSADLELERATDVKWDDQAEISGSSPILSITISEDGSMSIKNEEE
                     EQTLTILIRTGLLVISGLFPVSIPITAAAWYLWEVKKQRAGVLWDVPSPPPVGKAELE
                     DGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGKRIEPSWADVKKDLI
                     SYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFRTNTGTIGAVSLDFSPGTS
                     GSPIVDKKGKVVGLYGNGVVTRSGAYVSAIAQTEKSIEDNPEIEDDIFRKRRLTIMDL
                     HPGAGKTKRYLPAIVREAIKRGLRTLILAPTRVVAAEMEEALRGLPIRYQTPAIRAEH
                     TGREIVDLMCHATFTMRLLSPIRVPNYNLIIMDEAHFTDPASIAARGYISTRVEMGEA
                     AGIFMTATPPGSRDPFPQSNAPIMDEEREIPERSWNSGHEWVTDFKGKTVWFVPSIKT
                     GNDIAACLRKNGKRVIQLSRKTFDSEYVKTRTNDWDFVVTTDISEMGANFKAERVIDP
                     RRCMKPVILTDGEERVILAGPMPVTHSSAAQRRGRIGRNPRNENDQYIYMGEPLENDE
                     DCAHWKEAKMLLDNINTPEGIIPSLFEPEREKEVDAIDGEYRLRGEARKTFVDLMRRGD
                     LPVWLAYKVAAEGINYADRRWCFDGTRNNQILEENVEVEIWTKEGERKKLKPRWLDAR
                     IYSDPLALKEFKEFAAGRKSLTLNLITEMGRLPTFMTQKARDALDNLAVLHTAEAGGK
                     AYNHALSELPETLETLLLLTLLATVTGGIFLFLMSGRGMGKMTLGMCCIITASILLWY
                     AQIQPHWIAASIILEFFLIVLLIPEPEKQRTPQDNQLTYVIIAILTVVAATMANEMGF
                     LEKTKKDLGLGNIATQQPESNILDIDLRPASAWTLYAVATTFITPMLRHSIENSSVNV
                     SLTAIANQATVLMGLGKGWPLSKMDIGVPLLAIGCYSQVNPITLTAALLLLVAHYAII
                     GPGLQAKATREAQKRAAAGIMKNPTVDGITVIDLDPIPYDPKFEKQLGQVMLLVLCVT
                     QVLMMRTTWALCEALTLATGPVSTLWEGNPGRFWNTTIAVSMANIFRGSYLAGAGLLF
                     SIMKNTTSTRRGTGNIGETLGEKWKSRLNALGKSEFQIYKKSGIQEVDRTLAKEGIKR
                     GETDHHAVSRGSAKLRWFVERNLVTPEGKVVDLGCGRGGWSYYCGGLKNVREVKGLTK
                     GGPGHEEPIPMSTYGWNLVRLQSGVDVFFVPPEKCDTLLCDIGESSPNPTVEAGRTLR
                     VLNLVENWLNNNTQFCVKVLNPYMPSVIERMETLQRKYGGALVRNPLSRNSTHEMYWV
                     SNASGNIVSSVNMISRMLINRFTMRHKKATYEPDVDLGSGTRNIGIESETPNLDIIGK
                     RIEKIKQEHETSWHYDQDHPYKTWAYHGSYETKQTGSASSMVNGVVRLLTKPWDVVPM
                     VTQMAMTDTTPFGQQRVFKEKVDTRTQEPKEGTKKLMKITAEWLWKELGKKKTPRMCT
                     REEFTKKVRSNAALGAIFTDENKWKSAREAVEDSRFWELVDKERNLHLEGKCETCVYN
                     MMGKREKKLGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWFSRENSLSGVEGEG
                     LHKLGYILREVSKKEGGAMYADDTAGWDTRITIEDLKNEEMITNHMAGEHKKLAEAIF
                     KLTYQNKVVRVQRPTPRGTVMDIISRRDQRGSGQVGTYGLNTFTNMEAQLIRQMEGEG
                     IFKSIQHLTASEEIAVQDWLVRVGRERLSRMAISGDDCVVKPLDDRFARALTALNDMG
                     KVRKDIQQWEPSRGWNDWTQVPFCSHHFHELIMKDGRTLVVPCRNQDELIGRARISQG
                     AGWSLRETACLGKSYAQMWSLMYFHRRDLRLAANAICSAVPSHWIPTSRTTWSIHASH
                     EWMTTEDMLTVWNRVWILENPWMEDKTPVESWEEIPYLGKREDQWCGSLIGLTSRATW
                     AKNIQTAINQVRSLIGNEEYTDYMPSMKRFRREEEEAGVLW"
     mat_peptide     97. .438
                     /product="anchored capsid protein (anchC)"
     mat_peptide     97. .396
                     /product="virion capsid protein (virC)"
     mat_peptide     439. .936
                     /product="membrane precursor protein (prM)"
     mat_peptide     712. .936
                     /product="membrane protein (M)"
     mat_peptide     937. .2421
                     /product="envelope protein (E)"
     mat_peptide     2422. .3477
                     /product="NS1 protein"
     mat_peptide     3478. .4131
                     /product="NS2A protein"
     mat_peptide     4132. .4521
                     /product="NS2B protein"
     mat_peptide     4522. .6375
                     /product="NS3 protein"
     mat_peptide     6376. .6756
                     /product="NS4A protein"
     mat_peptide     6757. .6825
                     /product="2K protein"
     mat_peptide     6826. .7569
                     /product="NS4B protein"
     mat_peptide     7570. .10269
                     /product="NS5 protein"
     misc_feature    10650. .10658
                     /note="KpnI/AgeI linker"
     promoter        15111. .15129
                     /note="SP6"
ORIGIN

  Query Match            34.0%;  Score 1636.8;  DB 11;  Length 15129;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    |  |    |  ||||||||||||||||||||||||||||||
Db      11222 TGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11281

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11282 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11341

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11342 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11401

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11402 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11461

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11462 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11521

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11522 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11581

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11582 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11641

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      11642 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11701

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11702 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11761

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11762 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11821

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11822 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11881

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11882 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11941

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
```

```
Db      11942 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12001

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948

Db      12002 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12061

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008

Db      12062 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12121

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068

Db      12122 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12181

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128

Db      12182 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12241

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188

Db      12242 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12301

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248

Db      12302 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12361

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308

Db      12362 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12421

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368

Db      12422 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12481

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428

Db      12482 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12541

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488

Db      12542 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12601

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548

Db      12602 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12661

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608

Db      12662 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12721

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668

Db      12722 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12781

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728

Db      12782 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12841

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776

Db      12842 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12901

Qy       2777 ------------------------------------------------------------ 2776

Db      12902 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12961

Qy       2777 ------------------------------------------------------------ 2776

Db      12962 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13021

Qy       2777 ------------------------------------------------------------ 2776

Db      13022 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13081

Qy       2777 ------------------------------------------------------------ 2776

Db      13082 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13141

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805

Db      13142 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13201

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865

Db      13202 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13261

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925

Db      13262 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13321

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985

Db      13322 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13381

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045

Db      13382 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13441

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073

Db      13442 CTGGACGCGGATGAACAGGCAGACATCT 13469
```

```
RESULT 2
AY744150
LOCUS       AY744150               15129 bp    DNA     circular SYN 03-DEC-2004
DEFINITION  Dengue virus type 2 vector p4(delta30)-4995, complete sequence.
ACCESSION   AY744150
VERSION     AY744150.1  GI:56089728
KEYWORDS    .
SOURCE      Dengue virus type 2 vector p4(delta30)-4995
  ORGANISM  Dengue virus type 2 vector p4(delta30)-4995
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15129)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Hanley,K.A., Murphy,B.R. and
            Whitehead,S.S.
  TITLE     Vaccine candidates derived from a novel infectious cDNA clone of an
            American genotype dengue virus type 2
  JOURNAL   (er) BMC Infect. Dis. 4 (1), 39 (2004)
   PUBMED   15461822
REFERENCE   2  (bases 1 to 15129)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15129
                     /organism="Dengue virus type 2 vector p4(delta30)-4995"
                     /mol_type="other DNA"
                     /strain="Tonga/74"
                     /db_xref="taxon:301537"
                     /country="Tonga"
                     /focus
     source          1. .10683
                     /organism="Dengue virus type 2"
                     /mol_type="other DNA"
                     /db_xref="taxon:11060"
     source          10693. .15054
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     CDS             97. .10272
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAV70832.1"
```

```
                     /db_xref="GI:56089729"
                     /translation="MNNQRKKARNTPFNMLKRERNRVSTVQQLTKRFSLGMLQGRGPL
                     KLFMALVAFLRFLTIPPTAGILKRWGTIKKSKAINVLRGFRKEIGRMLNILNRRRRTV
                     GMIIMLTPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTKDGTNMCTLMAMDLGELCE
                     DTITYKCPFLKQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLE
                     TRTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRVLIFILLTAIA
                     PSMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPA
                     TLRKYCIEAKLTNTTTDSRCPTQGEPTLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGG
                     IVTCAMFTCKKNMEGKIVQPENLEYTVVITPHSGEEHAVGNDTGKHGKEVKITPQSSI
                     TEAELTGYGTVTMECSPRTGLDFNEMVLLQMEDKAWLVHRQWFLDLPLPWLPGADTQG
                     SNWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCR
                     LRMDKLQLKGMSYSMCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEK
                     RHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLDWFKKGSSIGQMF
                     ETTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIG
                     VIITWIGMNSRSTSLSVSLVLVGIVTLYLGVMVQADSGCVVSWKNKELKCGSGIFVTD
                     NVHTWTEQYKFQPESPSKLASAIQKAHEEGICGIRSVTRLENLMWKQITSELNHILSE
                     NEVKLTIMTGDIKGIMQVGKRSLRPQPTELRYSWKTWGKAKMLSTELHNQTFLIDGPE
                     TAECPNTNRAWNSLEVEDYGFGVFTTNIWLRLREKQDVFCDSKLMSAAIKDNRAVHAD
                     MGYWIESALNDTWKIEKASFIEVKSCHWPKSHTLWSNGVLESEMVIPKNFAGPVSQHN
                     NRPGYYTQTAGPWHLGKLEMDFDFCEGTTVVVTENCGNRGPSLRTTTASGKLITEWCC
                     RSCTLPPLRYRGEDGCWYGMEIRPLKEKEENLVSSLVTAGHGQIDNFSLGILGMALFL
                     EEMLRTRVGTKHAILLVAVSFVTLITGNMSFRDLGRVMVMVGATMTDDIGMGVTYLAL
                     LAAFRVRPTFAAGLLLRKLTSKELMMTTIGIVLLSQSSIPETILELTDALALGMMVLK
                     MVRNMEKYQLAVTIMAILCVPNAVILQNAWKVSCTILAVVSVSPLLLTSSQQKADWIP
                     LALTIKGLNPTAIFLTTLSRTNKKRSWPLNEAIMAVGMVSILASSLLKNDIPMTGPLV
                     AGGLLTVCYVLTGRSADLELERATDVKWDDQAEISGSSPILSITISEDGSMSIKNEEE
                     EQTLTILIRTGLLVISGLFPVSIPITAAAWYLWEVKKQRAGVLWDVPSPPPVGKAELE
                     DGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGKRIEPSWADVKKDLI
                     SYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFRTNTGTIGAVSLDFSPGTS
                     GSPIVDKKGKVVGLYGNGVVTRLGAYVSAIAQTEKSIEDNPEIEDDIFRKRRLTIMDL
                     HPGAGKTKRYLPAIVREAIKRGLRTLILAPTRVVAAEMEEALRGLPIRYQTPAIRAEH
                     TGREIVDLMCHATFTMRLLSPIRVPNYNLIIMDEAHFTDPASIAARGYISTRVEMGEA
                     AGIFMTATPPGSRDPFPQSNAPIMDEEREIPERSWNSGHEWVTDFKGKTVWFVPSIKT
                     GNDIAACLRKNGKRVIQLSRKTFDSEYVKTRTNDWDFVVTTDISEMGANFKAERVIDP
                     RRCMKPVILTDGEERVILAGPMPVTHSSAAQRRGRIGRNPRNENDQYIYMGEPLENDE
                     DCAHWKEAKMLLDNINTPEGIIPSLFEPEREKEVDAIDGEYRLRGEARKTFVDLMRRGD
                     LPVWLAYKVAAEGINYADRRWCFDGTRNNQILEENVEVEIWTKEGERKKLKPRWLDAR
                     IYSDPLALKEFKEFAAGRKSLTLNLITEMGRLPTFMTQKARDALDNLAVLHTAEAGGK
                     AYNHALSELPETLETLLLLTLLATVTGGIFLFLMSGRGMGKMTLGMCCIITASILLWY
                     AQIQPHWIAASIILEFFLIVLLIPEPEKQRTPQDNQLTYVIIAILTVVAATMANEMGF
                     LEKTKKDLGLGNIATQQPESNILDIDLRPASAWTLYAVATTFITPMLRHSIENSSVNV
                     SLTAIANQATVLMGLGKGWPLSKMDIGVPLLAIGCYSQVNPITLTAALLLLVAHYAII
                     GPGLQAKATREAQKRAAAGIMKNPTVDGITVIDLDPIPYDPKFEKQLGQVMLLVLCVT
                     QVLMMRTTWALCEALTLATGPVSTLWEGNPGRFWNTTIAVSMANIFRGSYLAGAGLLF
                     SIMKNTTSTRRGTGNIGETLGEKWKSRLNALGKSEFQIYKKSGIQEVDRTLAKEGIKR
                     GETDHHAVSRGSAKLRWFVERNLVTPEGKVVDLGCGRGGWSYYCGGLKNVREVKGLTK
                     GGPGHEEPIPMSTYGWNLVRLQSGVDVFFVPPEKCDTLLCDIGESSPNPTVEAGRTLR
                     VLNLVENWLNNNTQFCVKVLNPYMPSVIERMETLQRKYGGALVRNPLSRNSTHEMYWV
                     SNASGNIVSSVNMISRMLINRFTMRHKKATYEPDVDLGSGTRNIGIESETPNLDIIGK
                     RIEKIKQEHETSWHYDQDHPYKTWAYHGSYETKQTGSASSMVNGVVRLLTKPWDVVPM
                     VTQMAMTDTTPFGQQRVFKEKVDTRTQEPKEGTKKLMKITAEWLWKELGKKKTPRMCT
                     REEFTKKVRSNAALGAIFTDENKWKSAREAVEDSRFWELVDKERNLHLEGKCETCVYN
                     MMGKREKKLGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWFSRENSLSGVEGEG
                     LHKLGYILREVSKKEGGAMYADDTAGWDTRITIEDLKNEEMITNHMAGEHKKLAEAIF
                     KLTYQNKVVRVQRPTPRGTVMDIISRRDQRGSGQVGTYGLNTFTNMEAQLIRQMEGEG
                     IFKSIQHLTASEEIAVQDWLVRVGRERLSRMAISGDDCVVKPLDDRFARALTALNDMG
                     KVRKDIQQWEPSRGWNDWTQVPFCSHHFHELIMKDGRTLVVPCRNQDELIGRARISQG
                     AGWSLRETACLGKSYAQMWSLMYFHRRDLRLAANAICSAVPSHWIPTSRTTWSIHASH
                     EWMTTEDMLTVWNRVWILENPWMEDKTPVESWEEIPYLGKREDQWCGSLIGLTSRATW
                     AKNIQTAINQVRSLIGNEEYTDYMPSMKRFRREEEEAGVLW"
mat_peptide          97. .438
                     /product="anchored capsid protein (anchC)"
mat_peptide          97. .396
                     /product="virion capsid protein (virC)"
mat_peptide          439. .936
                     /product="membrane precursor protein (prM)"
mat_peptide          712. .936
                     /product="membrane protein (M)"
mat_peptide          937. .2421
                     /product="envelope protein (E)"
mat_peptide          2422. .3477
                     /product="NS1 protein"
mat_peptide          3478. .4131
                     /product="NS2A protein"
mat_peptide          4132. .4521
                     /product="NS2B protein"
mat_peptide          4522. .6375
                     /product="NS3 protein"
mat_peptide          6376. .6756
                     /product="NS4A protein"
mat_peptide          6757. .6825
                     /product="2K protein"
mat_peptide          6826. .7569
                     /product="NS4B protein"
mat_peptide          7570. .10269
                     /product="NS5 protein"
misc_feature         10684. .10692
                     /note="KpnI/AgeI linker"
promoter             15111. .15129
                     /note="SP6"
ORIGIN

  Query Match           34.0%;  Score 1636.8;  DB 11;  Length 15129;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    | |    | |||||||||||||||||||||||||||||||
Db      11222 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11281

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11282 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11341

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11342 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11401

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11402 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11461

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11462 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11521

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11522 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11581

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11582 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11641

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11642 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11701

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11702 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11761

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11762 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11821

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11822 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11881

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11882 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11941

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11942 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12001

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12002 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12061

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12062 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12121

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12122 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12181

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12182 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12241

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12242 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12301

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12302 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12361

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12362 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12421

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12422 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12481

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12482 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12541

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      12542 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12601

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12602 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12661

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12662 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12721

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12722 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12781

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12782 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12841

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG---------- 2776
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12842 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12901

Qy       2777 ------------------------------------------------------------ 2776

Db      12902 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12961

Qy       2777 ------------------------------------------------------------ 2776

Db      12962 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13021

Qy       2777 ------------------------------------------------------------ 2776

Db      13022 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13081

Qy       2777 ------------------------------------------------------------ 2776

Db      13082 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13141

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||
Db      13142 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13201

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13202 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13261

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13262 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13321

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13322 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13381

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13382 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13441

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
              ||||||||||||||||||||||  |||
Db      13442 CTGGACGCGGATGAACAGGCAGACATCT 13469
```

```
RESULT 3
AY656170
LOCUS       AY656170               15145 bp    DNA     circular SYN 14-JAN-2005
DEFINITION  Dengue virus type 3 vector p3(delta30), complete sequence.
ACCESSION   AY656170
VERSION     AY656170.1  GI:50345912
KEYWORDS    .
SOURCE      Dengue virus type 3 vector p3(delta30)
  ORGANISM  Dengue virus type 3 vector p3(delta30)
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15145)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Firestone,C.Y., Hanley,K.A.,
            Murphy,B.R. and Whitehead,S.S.
  TITLE     Genetically modified, live attenuated dengue virus type 3 vaccine
            candidates
  JOURNAL   Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
   PUBMED   15642976
REFERENCE   2  (bases 1 to 15145)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15145
                     /organism="Dengue virus type 3 vector p3(delta30)"
                     /mol_type="other DNA"
                     /db_xref="taxon:285000"
                     /focus
     source          1. .2345
                     /organism="Dengue virus type 3"
                     /mol_type="other DNA"
                     /db_xref="taxon:11069"
                     /note="Sleman/78"
     source          2369. .10699
                     /organism="Dengue virus type 3"
                     /mol_type="other DNA"
                     /db_xref="taxon:11069"
                     /note="Sleman/78"
     source          10709. .15070
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     source          15071. .15145
```

```
                     /organism="Cloning vector pGEM-3"
                     /mol_type="other DNA"
                     /db_xref="taxon:90108"
     misc_feature    95. .10287
                     /note="encodes polyprotein precursor once SpeI linker
                     sequence is removed prior to preparation of infectious
                     transcripts corresponding to the viral genome"
     misc_feature    2346. .2368
                     /note="SpeI linker; required for stable replication in
                     Escherichia coli"
     misc_feature    10700. .10708
                     /note="KpnI/AgeI linker"
     promoter        15127. .15145
                     /note="SP6"
ORIGIN

  Query Match             34.0%;  Score 1636.8;  DB 11;  Length 15145;
  Best Local Similarity   86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
Db      11238 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11297

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
Db      11298 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11357

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
Db      11358 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11417

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
Db      11418 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11477

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
Db      11478 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11537

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
Db      11538 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11597

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
Db      11598 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11657

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
Db      11658 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11717

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
Db      11718 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11777

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
Db      11778 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11837

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
Db      11838 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11897

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
Db      11898 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11957

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
Db      11958 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12017

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948

Db      12018 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12077

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
Db      12078 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12137

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
Db      12138 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12197

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
Db      12198 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12257

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
Db      12258 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12317

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
Db      12318 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12377

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
Db      12378 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12437

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
Db      12438 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12497

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
Db      12498 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12557

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
Db      12558 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12617

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
Db      12618 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12677

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
Db      12678 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12737

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
Db      12738 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12797

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
Db      12798 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12857

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
Db      12858 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12917

Qy       2777 ---------------------------------------------------------- 2776
Db      12918 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12977

Qy       2777 ---------------------------------------------------------- 2776
Db      12978 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13037

Qy       2777 ---------------------------------------------------------- 2776
Db      13038 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13097

Qy       2777 ---------------------------------------------------------- 2776
```

```
Db      13098 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13157

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                             ||||||||||||||||||||||||||||
Db      13158 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13217

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13218 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13277

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13278 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13337

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13338 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13397

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13398 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13457

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
              |||||||||||||||||||||  |||
Db      13458 CTGGACGCGGATGAACAGGCAGACATCT 13485
```

```
RESULT 4
AY744148
LOCUS       AY744148               15159 bp    DNA     circular SYN 03-DEC-2004
DEFINITION  Dengue virus type 2 vector p2, complete sequence.
ACCESSION   AY744148
VERSION     AY744148.1  GI:56089724
KEYWORDS    .
SOURCE      Dengue virus type 2 vector p2
  ORGANISM  Dengue virus type 2 vector p2
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15159)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Hanley,K.A., Murphy,B.R. and
            Whitehead,S.S.
  TITLE     Vaccine candidates derived from a novel infectious cDNA clone of an
            American genotype dengue virus type 2
  JOURNAL   (er) BMC Infect. Dis. 4 (1), 39 (2004)
   PUBMED   15461822
REFERENCE   2  (bases 1 to 15159)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15159
                     /organism="Dengue virus type 2 vector p2"
                     /mol_type="other DNA"
                     /strain="Tonga/74"
                     /db_xref="taxon:301535"
                     /country="Tonga"
                     /focus
     source          1. .10713
                     /organism="Dengue virus type 2"
                     /mol_type="other DNA"
                     /db_xref="taxon:11060"
     source          10723. .15084
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     CDS             97. .10272
                     /codon_start=1
                     /product="polyprotein"
                     /protein_id="AAV70830.1"
                     /db_xref="GI:56089725"
                     /translation="MNNQRKKARNTPFNMLKRERNRVSTVQQLTKRFSLGMLQGRGPL
                     KLFMALVAFLRFLTIPPTAGILKRWGTIKKSKAINVLRGFRKEIGRMLNILNRRRRTV
                     GMIIMLTPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTKDGTNMCTLMAMDLGELCE
                     DTITYKCPFLKQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLE
                     TRTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRVLIFILLTAIA
                     PSMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPA
                     TLRKYCIEAKLTNTTTDSRCPTQGEPTLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGG
                     IVTCAMFTCKKNMEGKIVQPENLEYTVVITPHSGEEHAVGNDTGKHGKEVKITPQSSI
                     TEAELTGYGTVTMECSPRTGLDFNEMVLLQMEDKAWLVHRQWFLDLPLPWLPGADTQG
                     SNWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCR
                     LRMDKLQLKGMSYSMCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEK
                     RHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLDWFKKGSSIGQMF
                     ETTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIG
                     VIITWIGMNSRSTSLSVSLVLVGIVTLYLGVMVQADSGCVVSWKNKELKCGSGIFVTD
                     NVHTWTEQYKFQPESPSKLASAIQKAHEEGICGIRSVTRLENLMWKQITSELNHILSE
                     NEVKLTIMTGDIKGIMQVGKRSLRPQPTELRYSWKTWGKAKMLSTELHNQTFLIDGPE
                     TAECPNTNRAWNSLEVEDYGFGVFTTNIWLRLREKQDVFCDSKLMSAAIKDNRAVHAD
                     MGYWIESALNDTWKIEKASFIEVKSCHWPKSHTLWSNGVLESEMVIPKNFAGPVSQHN
                     NRPGYYTQTAGPWHLGKLEMDFDFCEGTTVVVTENCGNRGPSLRTTTASGKLITEWCC
                     RSCTLPPLRYRGEDGCWYGMEIRPLKEKEENLVSSLVTAGHGQIDNFSLGILGMALFL
                     EEMLRTRVGTKHAILLVAVSFVTLITGNMSFRDLGRVMVMVGATMTDDIGMGVTYLAL
                     LAAFRVRPTFAAGLLLRKLTSKELMMTTIGIVLLSQSSIPETILELTDALALGMMVLK
                     MVRNMEKYQLAVTIMAILCVPNAVILQNAWKVSCTILAVVSVSPLLLTSSQQKADWIP
                     LALTIKGLNPTAIFLTTLSRTNKKRSWPLNEAIMAVGMVSILASSLLKNDIPMTGPLV
                     AGGLLTVCYVLTGRSADLELERATDVKWDDQAEISGSSPILSITISEDGSMSIKNEEE
                     EQTLTILIRTGLLVISGLFPVSIPITAAAWYLWEVKKQRAGVLWDVPSPPPVGKAELE
                     DGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGKRIEPSWADVKKDLI
                     SYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFRTNTGTIGAVSLDFSPGTS
                     GSPIVDKKGKVVGLYGNGVVTRSGAYVSAIAQTEKSIEDNPEIEDDIFRKRRLTIMDL
                     HPGAGKTKRYLPAIVREAIKRGLRTLILAPTRVVAAEMEEALRGLPIRYQTPAIRAEH
                     TGREIVDLMCHATFTMRLLSPIRVPNYNLIIMDEAHFTDPASIAARGYISTRVEMGEA
                     AGIFMTATPPGSRDPFPQSNAPIMDEEREIPERSWNSGHEWVTDFKGKTVWFVPSIKT
                     GNDIAACLRKNGKRVIQLSRKTFDSEYVKTRTNDWDFVVTTDISEMGANFKAERVIDP
                     RRCMKPVILTDGEERVILAGPMPVTHSSAAQRRGRIGRNPRNENDQYIYMGEPLENDE
                     DCAHWKEAKMLLDNINTPEGIIPSLFEPEREKEVDAIDGEYRLRGEARKTFVDLMRRGD
                     LPVWLAYKVAAEGINYADRRWCFDGTRNNQILEENVEVEIWTKEGERKKLKPRWLDAR
                     IYSDPLALKEFKEFAAGRKSLTLNLITEMGRLPTFMTQKARDALDNLAVLHTAEAGGK
                     AYNHALSELPETLETLLLLTLLATVTGGIFLFLMSGRGMGKMTLGMCCIITASILLWY
                     AQIQPHWIAASIILEFFLIVLLIPEPEKQRTPQDNQLTYVIIAILTVVAATMANEMGF
                     LEKTKKDLGLGNIATQQPESNILDIDLRPASAWTLYAVATTFITPMLRHSIENSSVNV
                     SLTAIANQATVLMGLGKGWPLSKMDIGVPLLAIGCYSQVNPITLTAALLLLVAHYAII
                     GPGLQAKATREAQKRAAAGIMKNPTVDGITVIDLDPIPYDPKFEKQLGQVMLLVLCVT
                     QVLMMRTTWALCEALTLATGPVSTLWEGNPGRFWNTTIAVSMANIFRGSYLAGAGLLF
                     SIMKNTTSTRRGTGNIGETLGEKWKSRLNALGKSEFQIYKKSGIQEVDRTLAKEGIKR
                     GETDHHAVSRGSAKLRWFVERNLVTPEGKVVDLGCGRGGWSYYCGGLKNVREVKGLTK
                     GGPGHEEPIPMSTYGWNLVRLQSGVDVFFVPPEKCDTLLCDIGESSPNPTVEAGRTLR
                     VLNLVENWLNNNTQFCVKVLNPYMPSVIERMETLQRKYGGALVRNPLSRNSTHEMYWV
                     SNASGNIVSSVNMISRMLINRFTMRHKKATYEPDVDLGSGTRNIGIESETPNLDIIGK
                     RIEKIKQEHETSWHYDQDHPYKTWAYHGSYETKQTGSASSMVNGVVRLLTKPWDVVPM
                     VTQMAMTDTTPFGQQRVFKEKVDTRTQEPKEGTKKLMKITAEWLWKELGKKKTPRMCT
                     REEFTKKVRSNAALGAIFTDENKWKSAREAVEDSRFWELVDKERNLHLEGKCETCVYN
                     MMGKREKKLGEFGKAKGSRAIWYMWLGARFLEFEALGFLNEDHWFSRENSLSGVEGEG
                     LHKLGYILREVSKKEGGAMYADDTAGWDTRITIEDLKNEEMITNHMAGEHKKLAEAIF
                     KLTYQNKVVRVQRPTPRGTVMDIISRRDQRGSGQVGTYGLNTFTNMEAQLIRQMEGEG
                     IFKSIQHLTASEEIAVQDWLVRVGRERLSRMAISGDDCVVKPLDDRFARALTALNDMG
                     KVRKDIQQWEPSRGWNDWTQVPFCSHHFHELIMKDGRTLVVPCRNQDELIGRARISQG
                     AGWSLRETACLGKSYAQMWSLMYFHRRDLRLAANAICSAVPSHWIPTSRTTWSIHASH
                     EWMTTEDMLTVWNRVWILENPWMEDKTPVESWEEIPYLGKREDQWCGSLIGLTSRATW
                     AKNIQTAINQVRSLIGNEEYTDYMPSMKRFRREEEEAGVLW"
     mat_peptide     97. .438
                     /product="anchored capsid protein (anchC)"
     mat_peptide     97. .396
                     /product="virion capsid protein (virC)"
     mat_peptide     439. .936
                     /product="membrane precursor protein (prM)"
     mat_peptide     712. .936
                     /product="membrane protein (M)"
     mat_peptide     937. .2421
                     /product="envelope protein (E)"
     mat_peptide     2422. .3477
                     /product="NS1 protein"
     mat_peptide     3478. .4131
                     /product="NS2A protein"
     mat_peptide     4132. .4521
                     /product="NS2B protein"
     mat_peptide     4522. .6375
```

```
                     /product="NS3 protein"
     mat_peptide     6376. .6756
                     /product="NS4A protein"
     mat_peptide     6757. .6825
                     /product="2K protein"
     mat_peptide     6826. .7569
                     /product="NS4B protein"
     mat_peptide     7570. .10269
                     /product="NS5 protein"
     misc_feature    10714. .10722
                     /note="KpnI/AgeI linker"
     promoter        15141. .15159
                     /note="SP6"
ORIGIN
```

Query Match 34.0%; Score 1636.8; DB 11; Length 15159;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1;

```
Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
Db      11252 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11311

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
Db      11312 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11371

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
Db      11372 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11431

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
Db      11432 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11491

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
Db      11492 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11551

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
Db      11552 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11611

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
Db      11612 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11671

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
Db      11672 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11731

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
Db      11732 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11791

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
Db      11792 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11851

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
Db      11852 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11911

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
Db      11912 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11971

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
Db      11972 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12031

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
Db      12032 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12091

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
Db      12092 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12151

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
Db      12152 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12211

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
Db      12212 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12271

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
Db      12272 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12331

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
Db      12332 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12391

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
Db      12392 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12451

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
Db      12452 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12511

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
Db      12512 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12571

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
Db      12572 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12631

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
Db      12632 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12691

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
Db      12692 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12751

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
Db      12752 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12811

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
Db      12812 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12871

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
Db      12872 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12931

Qy       2777 ---------------------------------------------------------- 2776
Db      12932 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12991

Qy       2777 ---------------------------------------------------------- 2776
Db      12992 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13051

Qy       2777 ---------------------------------------------------------- 2776
Db      13052 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13111

Qy       2777 ---------------------------------------------------------- 2776
Db      13112 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13171
```

```
Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||||
Db      13172 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13231

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13232 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13291

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13292 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13351

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13352 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13411

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13412 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13471

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
              |||||||||||||||||||||||| |||
Db      13472 CTGGACGCGGATGAACAGGCAGACATCT 13499
```

```
RESULT 5
AY656169
LOCUS       AY656169               15176 bp    DNA     circular SYN 14-JAN-2005
DEFINITION  Dengue virus type 3 vector p3, complete sequence.
ACCESSION   AY656169
VERSION     AY656169.1  GI:50345911
KEYWORDS    .
SOURCE      Dengue virus type 3 vector p3
  ORGANISM  Dengue virus type 3 vector p3
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15176)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Firestone,C.Y., Hanley,K.A.,
            Murphy,B.R. and Whitehead,S.S.
  TITLE     Genetically modified, live attenuated dengue virus type 3 vaccine
            candidates
  JOURNAL   Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
   PUBMED   15642976
REFERENCE   2  (bases 1 to 15176)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15176
                     /organism="Dengue virus type 3 vector p3"
                     /mol_type="other DNA"
                     /db_xref="taxon:284999"
                     /focus
     source          1. .2345
                     /organism="Dengue virus type 3"
                     /mol_type="other DNA"
                     /db_xref="taxon:11069"
                     /note="Sleman/78"
     source          2369. .10730
                     /organism="Dengue virus type 3"
                     /mol_type="other DNA"
                     /db_xref="taxon:11069"
                     /note="Sleman/78"
     source          10740. .15101
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     source          15102. .15176
                     /organism="Cloning vector pGEM-3"
                     /mol_type="other DNA"
                     /db_xref="taxon:90108"
     misc_feature    95. .10287
                     /note="encodes polyprotein precursor once SpeI linker
                     sequence is removed prior to preparation of infectious
                     transcripts corresponding to the viral genome"
     misc_feature    2346. .2368
                     /note="SpeI linker; required for stable replication in
                     Escherichia coli"
     misc_feature    10731. .10739
                     /note="KpnI/AgeI linker"
     promoter        15158. .15176
                     /note="SP6"
ORIGIN
```

```
  Query Match             34.0%;  Score 1636.8;  DB 11;  Length 15176;
  Best Local Similarity   86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
              ||| ||| ||  | |    | |      | ||||||||||||||||||||||||||||||
Db      11269 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11328

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11329 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11388

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11389 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11448

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11449 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11508

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11509 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11568

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11569 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11628

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11629 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11688

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      11689 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11748

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11749 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11808

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11809 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11868

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11869 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11928

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11929 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11988

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11989 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12048

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12049 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12108

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
```

```
Db      12109 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12168
Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
Db      12169 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12228
Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
Db      12229 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12288
Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
Db      12289 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12348
Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
Db      12349 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12408
Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
Db      12409 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12468
Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
Db      12469 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12528
Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
Db      12529 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12588
Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
Db      12589 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12648
Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
Db      12649 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12708
Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
Db      12709 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12768
Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
Db      12769 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12828
Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
Db      12829 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12888
Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG---------- 2776
Db      12889 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12948
Qy       2777 ---------------------------------------------------------- 2776
Db      12949 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 13008
Qy       2777 ---------------------------------------------------------- 2776
Db      13009 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13068
Qy       2777 ---------------------------------------------------------- 2776
Db      13069 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13128
Qy       2777 ---------------------------------------------------------- 2776
Db      13129 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13188
Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
Db      13189 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13248
Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
Db      13249 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13308
Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
Db      13309 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13368
Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
Db      13369 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13428
Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
Db      13429 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13488
Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
Db      13489 CTGGACGCGGATGAACAGGCAGACATCT 13516
```

```
RESULT 6
AY243469
LOCUS       AY243469               15237 bp    DNA     circular SYN 20-OCT-2003
DEFINITION  Chimeric Dengue virus vector p4(Delta30)-D2-CME, complete sequence.
ACCESSION   AY243469
VERSION     AY243469.1  GI:30026605
KEYWORDS    .
SOURCE      Chimeric Dengue virus vector p4(Delta30)-D2-CME
  ORGANISM  Chimeric Dengue virus vector p4(Delta30)-D2-CME
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15237)
  AUTHORS   Whitehead,S.S., Hanley,K.A., Blaney,J.E. Jr., Gilmore,L.E.,
            Elkins,W.R. and Murphy,B.R.
  TITLE     Substitution of the structural genes of dengue virus type 4 with
            those of type 2 results in chimeric vaccine candidates which are
            attenuated for mosquitoes, mice, and rhesus monkeys
  JOURNAL   Vaccine 21 (27-28), 4307-4316 (2003)
   PUBMED   14505913
REFERENCE   2  (bases 1 to 15237)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2003) LID, NIAID, 50 South Dr., Room 6515,
            Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15237
                     /organism="Chimeric Dengue virus vector
                     p4(Delta30)-D2-CME"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:227158"
                     /focus
     source          1. .88
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     source          89. .2343
                     /organism="Dengue virus type 2 (NGC-prototype)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11065"
     source          2344. .10616
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     CDS             97. .10263
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAP06255.1"
                     /db_xref="GI:30026606"
                     /translation="MNNQRKKARNTPFNMLKRERNRVSTVQQLTKRFSLGMLQGRGPL
                     KLFMALVAFLRFLTIPPTAGILKRWGTIKKSKAINVLRGFRKEIGRMLNILNRRRRTA
                     GMIIMLIPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTEDGVNMCTLMAMDLGELCE
```

```
DTITYKCPLLRQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLE
TRTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRALIFILLTAVA
PSMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPA
TLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGG
IVTCAMFTCKKNMEGKVVQPENLEYTIVITPHSGEEHAVGNDTGKHGKEIKITPQSSI
TEAELTGYGTVTMECSPRTGLDFNEMVLLQMENKAWLVHRQWFLDLPLPWLPGADTQG
SNWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCR
LRMDKLQLKGMSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEK
RHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMF
ETTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIG
VIITWIGMNSRNTSMAMTCIAVGGITLFLGFTVQADMGCVASWSGKELKCGSGIFVVD
NVHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLENVMWKQITNELNYVLWE
GGHDLTVVAGDVKGVLTKGKRALTPPVSDLKYSWKTWGKAKIFTPEARNSTFLIDGPD
TSECPNERRAWNSLEVEDYGFGMFTTNIWMKFREGSSEVCDHRLMSAAIKDQKAVHAD
MGYWIESSKNQTWQIEKASLIEVKTCLWPKTHTLWSNGVLESQMLIPKSYAGPFSQHN
YRQGYATQTVGPWHLGKLEIDFGECPGTTVTIQEDCDHRGPSLRTTTASGKLVTQWCC
RSCTMPPLRFLGEDGCWYGMEIRPLSEKEENMVKSQVTAGQGTSETFSMGLLCLTLFV
EECLRRRVTRKHMILVVVITLCAIILGGLTWMDLLRALIMLGDTMSGRIGGQIHLAIM
AVFKMSPGYVLGVFLRKLTSRETALMVIGMAMTTVLSIPHDLMELIDGISLGLILLKI
VTQFDNTQVGTLALSLTFIRSTMPLVMAWRTIMAVLFVVTLIPLCRTSCLQKQSHWVE
ITALILGAQALPVYLMTLMKGASRRSWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMV
AGGLLLAAYVMSGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEE
TNMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQRSGALWDVPSPAATKKAALS
EGVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMI
SYGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTS
GSPIINRKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVDEDIFRKKRLTIMDL
HPGAGKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEH
TGREIVDLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEA
AAIFMTATPPGATDPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKA
GNDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDP
RRCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDE
DHAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGD
LPVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDAR
VYADPMALKDFKEFASGRKSITLDILTEIASLPTYLSSRAKLALDNIVMLHTTERGGR
AYQHALNELPESLETLMLVALLGAMTAGIFLFFMQGKGIGKLSMGLITIAVASGLLWV
AEIQPQWIAASIILEFFLMVLLIPEPEKQRTPQDNQLIYVILTILTIIGLIAANEMGL
IEKTKTDFGFYQVKTETTILDVDLRPASAWTLYAVATTILTPMLRHTIENTSANLSLA
AIANQAAVLMGLGKGWPLHRMDLGVPLLAMGCYSQVNPTTLTASLVMLLVHYAIIGPG
LQAKATREAQKRTAAGIMKNPTVDGITVIDLEPISYDPKFEKQLGQVMLLVLCAGQLL
LMRTTWAFCEVLTLATGPILTLWEGNPGRFWNTTIAVSTANIFRGSYLAGAGLAFSLI
KNAQTPRRGTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSK
IKHAVSRGSSKIRWIVERGMVKPKGKVVDLGCGRGGWSYYMATLKNVTEVKGYTKGGP
GHEEPIPMATYGWNLVKLHSGVDVFYKPTEQVDTLLCDIGESSSNPTIEEGRTLRVLK
MVEPWLSSKPEFCIKVLNPYMPTVIEELEKLQRKHGGNLVRCPLSRNSTHEMYWVSGA
SGNIVSSVNTTSKMLLNRFTTRHRKPTYEKDVDLGAGTRSVSTETEKPDMTIIGRRLQ
RLQEEHKETWHYDQENPYRTWAYHGSYEAPSTGSASSMVNGVVKLLTKPWDVIPMVTQ
LAMTDTTPFGQQRVFKEKVDTRTPQPKPGTRMVMTTTANWLWALLGKKKNPRLCTREE
FISKVRSNAAIGAVFQEEQGWTSASEAVNDSRFWELVDKERALHQEGKCESCVYNMMG
KREKKLGEFGRAKGSRAIWYMWLGARFLEFEALGFLNEDHWFGRENSWSGVEGEGLHR
LGYILEEIDKKDGDLMYADDTAGWDTRITEDDLQNEELITEQMAPHHKILAKAIFKLT
YQNKVVKVLRPTPRGAVMDIISRKDQRGSGQVGTYGLNTFTNMEVQLIRQMEAEGVIT
QDDMQNPKGLKERVEKWLKECGVDRLKRMAISGDDCVVKPLDERFGTSLLFLNDMGKV
RKDIPQWEPSKGWKNWQEVPFCSHHFHKIFMKDGRSLVVPCRNQDELIGRARISQGAG
WSLRETACLGKAYAQMWSLMYFHRRDLRLASMAICSAVPTEWFPTSRTTWSIHAHHQW
MTTEDMLKVWNRVWIEDNPNMTDKTPVHSWEDIPYLGKREDLWCGSLIGLSSRATWAK
NIHTAITQVRNLIGKEEEYVDYMPVMKRYSAPSESEGVL"
mat_peptide 97. .438
/product="anchored capsid protein"
/note="anchC"
mat_peptide 97. .396
/product="virion capsid protein"
/note="virC"
mat_peptide 439. .936
/product="membrane precursor protein"
/note="prM"
mat_peptide 712. .936
/product="membrane protein"
/note="M"
mat_peptide 937. .2421
/product="envelope protein"
/note="E"
mat_peptide 2422. .3477
/product="NS1 protein"
mat_peptide 3478. .4131
/product="NS2A protein"
mat_peptide 4132. .4521
/product="NS2B protein"
mat_peptide 4522. .6375
/product="NS3 protein"
mat_peptide 6376. .6756
/product="NS4A protein"
mat_peptide 6757. .6825
/product="2K protein"
mat_peptide 6826. .7560
/product="NS4B protein"
mat_peptide 7561. .10260
/product="NS5 protein"
ORIGIN

Query Match 34.0%; Score 1636.8; DB 11; Length 15237;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1;

Qy 1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
Db 11155 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11214

Qy 1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
Db 11215 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11274

Qy 1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
Db 11275 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11334

Qy 1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
Db 11335 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11394

Qy 1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
Db 11395 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11454

Qy 1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
Db 11455 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11514

Qy 1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
Db 11515 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11574

Qy 1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
Db 11575 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11634

Qy 1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
Db 11635 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11694

Qy 1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
Db 11695 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11754

Qy 1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
Db 11755 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11814

Qy 1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
Db 11815 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11874

Qy 1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
Db 11875 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11934

Qy 1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
```

```
Db      11935 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 11994
Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
Db      11995 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12054
Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
Db      12055 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12114
Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
Db      12115 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12174
Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
Db      12175 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12234
Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
Db      12235 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12294
Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
Db      12295 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12354
Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
Db      12355 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12414
Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
Db      12415 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12474
Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
Db      12475 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12534
Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
Db      12535 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12594
Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
Db      12595 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12654
Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
Db      12655 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12714
Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
Db      12715 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12774
Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
Db      12775 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12834
Qy       2777 ------------------------------------------------------------ 2776
Db      12835 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12894
Qy       2777 ------------------------------------------------------------ 2776
Db      12895 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12954
Qy       2777 ------------------------------------------------------------ 2776
Db      12955 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13014
Qy       2777 ------------------------------------------------------------ 2776
```

```
Db      13015 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13074
Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
Db      13075 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13134
Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
Db      13135 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13194
Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
Db      13195 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13254
Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
Db      13255 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13314
Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
Db      13315 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13374
Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
Db      13375 CTGGACGCGGATGAACAGGCAGACATCT 13402
```

```
RESULT 7
AY243467
LOCUS       AY243467               15239 bp    DNA     circular SYN 20-OCT-2003
DEFINITION  Chimeric Dengue virus vector p4(Delta30)-D2-ME, complete sequence.
ACCESSION   AY243467
VERSION     AY243467.1  GI:30026601
KEYWORDS    .
SOURCE      Chimeric Dengue virus vector p4(Delta30)-D2-ME
  ORGANISM  Chimeric Dengue virus vector p4(Delta30)-D2-ME
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15239)
  AUTHORS   Whitehead,S.S., Hanley,K.A., Blaney,J.E. Jr., Gilmore,L.E.,
            Elkins,W.R. and Murphy,B.R.
  TITLE     Substitution of the structural genes of dengue virus type 4 with
            those of type 2 results in chimeric vaccine candidates which are
            attenuated for mosquitoes, mice, and rhesus monkeys
  JOURNAL   Vaccine 21 (27-28), 4307-4316 (2003)
   PUBMED   14505913
REFERENCE   2  (bases 1 to 15239)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2003) LID, NIAID, 50 South Dr., Room 6515,
            Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15239
                     /organism="Chimeric Dengue virus vector p4(Delta30)-D2-ME"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:227156"
                     /focus
     source          1. .404
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     source          405. .2345
                     /organism="Dengue virus type 2 (NGC-prototype)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11065"
     source          2346. .10618
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     CDS             102. .10265
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAP06253.1"
                     /db_xref="GI:30026602"
```

```
                     /translation="MNQRKKVVRPPFNMLKRERNRVSTPQGLVKRFSTGLFSGKGPLR
                     MVLAFITFLRVLSIPPTAGILKRWGQLKKNKAIKILIGFRKEIGRMLNILNGRKRSAG
                     MIIMLIPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTEDGVNMCTLMAMDLGELCED
                     TITYKCPLLRQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLET
                     RTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRALIFILLTAVAP
                     SMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPAT
                     LRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGGI
                     VTCAMFTCKKNMEGKVVQPENLEYTIVITPHSGEEHAVGNDTGKHGKEIKITPQSSIT
                     EAELTGYGTVTMECSPRTGLDFNEMVLLQMENKAWLVHRQWFLDLPLPWLPGADTQGS
                     NWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCRL
                     RMDKLQLKGMSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKR
                     HVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFE
                     TTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIGV
                     IITWIGMNSRNTSMAMTCIAVGGITLFLGFTVQADMGCVASWSGKELKCGSGIFVVDN
                     VHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLENVMWKQITNELNYVLWEG
                     GHDLTVVAGDVKGVLTKGKRALTPPVSDLKYSWKTWGKAKIFTPEARNSTFLIDGPDT
                     SECPNERRAWNSLEVEDYGFGMFTTNIWMKFREGSSEVCDHRLMSAAIKDQKAVHADM
                     GYWIESSKNQTWQIEKASLIEVKTCLWPKTHTLWSNGVLESQMLIPKSYAGPFSQHNY
                     RQGYATQTVGPWHLGKLEIDFGECPGTTVTIQEDCDHRGPSLRTTTASGKLVTQWCCR
                     SCTMPPLRFLGEDGCWYGMEIRPLSEKEENMVKSQVTAGQGTSETFSMGLLCLTLFVE
                     ECLRRRVTRKHMILVVVITLCAIILGGLTWMDLLRALIMLGDTMSGRIGGQIHLAIMA
                     VFKMSPGYVLGVFLRKLTSRETALMVIGMAMTTVLSIPHDLMELIDGISLGLILLKIV
                     TQFDNTQVGTLALSLTFIRSTMPLVMAWRTIMAVLFVVTLIPLCRTSCLQKQSHWVEI
                     TALILGAQALPVYLMTLMKGASRRSWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMVA
                     GGLLLAAYVMSGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEET
                     NMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQRSGALWDVPSPAATKKAALSE
                     GVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMIS
                     YGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSG
                     SPIINRKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVDEDIFRKKRLTIMDLH
                     PGAGKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEHT
                     GREIVDLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEAA
                     AIFMTATPPGATDPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAG
                     NDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPR
                     RCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDED
                     HAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGDL
                     PVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARV
                     YADPMALKDFKEFASGRKSITLDILTEIASLPTYLSSRAKLALDNIVMLHTTERGGRA
                     YQHALNELPESLETLMLVALLGAMTAGIFLFFMQGKGIGKLSMGLITIAVASGLLWVA
                     EIQPQWIAASIILEFFLMVLLIPEPEKQRTPQDNQLIYVILTILTIIGLIAANEMGLI
                     EKTKTDFGFYQVKTETTILDVDLRPASAWTLYAVATTILTPMLRHTIENTSANLSLAA
                     IANQAAVLMGLGKGWPLHRMDLGVPLLAMGCYSQVNPTTLTASLVMLLVHYAIIGPGL
                     QAKATREAQKRTAAGIMKNPTVDGITVIDLEPISYDPKFEKQLGQVMLLVLCAGQLLL
                     MRTTWAFCEVLTLATGPILTLWEGNPGRFWNTTIAVSTANIFRGSYLAGAGLAFSLIK
                     NAQTPRRGTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSKI
                     KHAVSRGSSKIRWIVERGMVKPKGKVVDLGCGRGGWSYYMATLKNVTEVKGYTKGGPG
                     HEEPIPMATYGWNLVKLHSGVDVFYKPTEQVDTLLCDIGESSSNPTIEEGRTLRVLKM
                     VEPWLSSKPEFCIKVLNPYMPTVIEELEKLQRKHGGNLVRCPLSRNSTHEMYWVSGAS
                     GNIVSSVNTTSKMLLNRFTTRHRKPTYEKDVDLGAGTRSVSTETEKPDMTIIGRRLQR
                     LQEEHKETWHYDQENPYRTWAYHGSYEAPSTGSASSMVNGVVKLLTKPWDVIPMVTQL
                     AMTDTTPFGQQRVFKEKVDTRTPQPKPGTRMVMTTTANWLWALLGKKKNPRLCTREEF
                     ISKVRSNAAIGAVFQEEQGWTSASEAVNDSRFWELVDKERALHQEGKCESCVYNMMGK
                     REKKLGEFGRAKGSRAIWYMWLGARFLEFEALGFLNEDHWFGRENSWSGVEGEGLHRL
                     GYILEEIDKKDGDLMYADDTAGWDTRITEDDLQNEELITEQMAPHHKILAKAIFKLTY
                     QNKVVKVLRPTPRGAVMDIISRKDQRGSGQVGTYGLNTFTNMEVQLIRQMEAEGVITQ
                     DDMQNPKGLKERVEKWLKECGVDRLKRMAISGDDCVVKPLDERFGTSLLFLNDMGKVR
                     KDIPQWEPSKGWKNWQEVPFCSHHFHKIFMKDGRSLVVPCRNQDELIGRARISQGAGW
                     SLRETACLGKAYAQMWSLMYFHRRDLRLASMAICSAVPTEWFPTSRTTWSIHAHHQWM
                     TTEDMLKVWNRVWIEDNPNMTDKTPVHSWEDIPYLGKREDLWCGSLIGLSSRATWAKN
                     IHTAITQVRNLIGKEEYVDYMPVMKRYSAPSESEGVL"
     mat_peptide     102. .440
                     /product="anchored capsid protein"
                     /note="anchC"
     mat_peptide     102. .398
                     /product="virion capsid protein"
                     /note="virC"
     mat_peptide     441. .938
                     /product="membrane precursor protein"
                     /note="prM"
     mat_peptide     714. .938
                     /product="membrane protein"
                     /note="M"
     mat_peptide     939. .2423
                     /product="envelope protein"
                     /note="E"
     mat_peptide     2424. .3479
                     /product="NS1 protein"
     mat_peptide     3480. .4133
                     /product="NS2A protein"
     mat_peptide     4134. .4523
                     /product="NS2B protein"
     mat_peptide     4524. .6377
                     /product="NS3 protein"
     mat_peptide     6378. .6758
                     /product="NS4A protein"
     mat_peptide     6759. .6827
                     /product="2K protein"
     mat_peptide     6828. .7562
                     /product="NS4B protein"
     mat_peptide     7563. .10262
                     /product="NS5 protein"
ORIGIN

  Query Match           34.0%;  Score 1636.8;  DB 11;  Length 15239;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| || | |  |  |  |    | |||||||||||||||||||||||||||||||
Db      11157 TGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11216

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11217 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11276

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11277 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11336

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11337 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11396

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11397 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11456

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11457 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11516

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11517 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11576

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11577 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11636

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11637 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11696

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11697 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11756

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11757 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11816

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11817 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11876

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11877 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11936
Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
Db      11937 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 11996
Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
Db      11997 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12056
Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
Db      12057 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12116
Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
Db      12117 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12176
Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
Db      12177 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12236
Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
Db      12237 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12296
Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
Db      12297 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12356
Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
Db      12357 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12416
Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
Db      12417 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12476
Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
Db      12477 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12536
Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
Db      12537 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12596
Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
Db      12597 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12656
Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
Db      12657 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12716
Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
Db      12717 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12776
Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
Db      12777 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12836
Qy       2777 ------------------------------------------------------------ 2776
Db      12837 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12896
Qy       2777 ------------------------------------------------------------ 2776
Db      12897 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12956
Qy       2777 ------------------------------------------------------------ 2776
Db      12957 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13016
```

```
Qy       2777 ------------------------------------------------------------ 2776
Db      13017 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13076
Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
Db      13077 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13136
Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
Db      13137 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13196
Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
Db      13197 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13256
Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
Db      13257 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13316
Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
Db      13317 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13376
Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
Db      13377 CTGGACGCGGATGAACAGGCAGACATCT 13404
```

```
RESULT 8
AY376438
LOCUS       AY376438               15239 bp    DNA     circular SYN 09-OCT-2003
DEFINITION  Dengue virus vector p4(Delta30), complete sequence.
ACCESSION   AY376438
VERSION     AY376438.1  GI:36312945
KEYWORDS    .
SOURCE      Dengue virus vector p4(Delta30)
  ORGANISM  Dengue virus vector p4(Delta30)
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15239)
  AUTHORS   Durbin,A.P., Karron,R.A., Sun,W., Vaughn,D.W., Reynolds,M.J.,
            Perreault,J.R., Thumar,B., Men,R., Lai,C.J., Elkins,W.R.,
            Chanock,R.M., Murphy,B.R. and Whitehead,S.S.
  TITLE     Attenuation and immunogenicity in humans of a live dengue virus
            type-4 vaccine candidate with a 30 nucleotide deletion in its
            3'-untranslated region
  JOURNAL   Am. J. Trop. Med. Hyg. 65 (5), 405-413 (2001)
   PUBMED   11716091
REFERENCE   2  (bases 1 to 15239)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2003) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15239
                     /organism="Dengue virus vector p4(Delta30)"
                     /mol_type="other DNA"
                     /db_xref="taxon:246753"
                     /focus
     source          1. .10618
                     /organism="Dengue virus type 4"
                     /mol_type="other DNA"
                     /db_xref="taxon:11070"
     source          10628. .14989
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     source          14990. .15239
                     /organism="Cloning vector pGEM-3"
                     /mol_type="other DNA"
                     /db_xref="taxon:90108"
     CDS             102. .10265
```

```
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAQ85129.1"
                     /db_xref="GI:36312946"
                     /translation="MNQRKKVVRPPFNMLKRERNRVSTPQGLVKRFSTGLFSGKGPLR
                     MVLAFITFLRVLSIPPTAGILKRWGQLKKNKAIKILIGFRKEIGRMLNILNGRKRSTI
                     TLLCLIPTVMAFSLSTRDGEPLMIVAKHERGRPLLFKTTEGINKCTLIAMDLGEMCED
                     TVTYKCPLLVNTEPEDIDCWCNLTSTWVMYGTCTQSGERRREKRSVALTPHSGMGLET
                     RAETWMSSEGAWKHAQRVESWILRNPGFALLAGFMAYMIGQTGIQRTVFFVLMMLVAP
                     SYGMRCVGVGNRDFVEGVSGGAWVDLVLEHGGCVTTMAQGKPTLDFELTKTTAKEVAL
                     LRTYCIEASISNITTATRCPTQGEPYLKEEQDQQYICRRDVVDRGWGNGCGLFGKGGV
                     VTCAKFSCSGKITGNLVQIENLEYTVVVTVHNGDTHAVGNDTSNHGVTAMITPRSPSV
                     EVKLPDYGELTLDCEPRSGIDFNEMILMKMKKKTWLVHKQWFLDLPLPWTAGADTSEV
                     HWNYKERMVTFKVPHAKRQDVTVLGSQEGAMHSALAGATEVDSGDGNHMFAGHLKCKV
                     RMEKLRIKGMSYTMCSGKFSIDKEMAETQHGTTVVKVKYEGAGAPCKVPIEIRDVNKE
                     KVVGRIISSTPLAENTNSVTNIELEPPFGDSYIVIGVGNSALTLHWFRKGSSIGKMFE
                     STYRGAKRMAILGETAWDFGSVGGLFTSLGKAVHQVFGSVYTTMFGGVSWMIRILIGF
                     LVLWIGTNSRNTSMAMTCIAVGGITLFLGFTVQADMGCVASWSGKELKCGSGIFVVDN
                     VHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLENVMWKQITNELNYVLWEG
                     GHDLTVVAGDVKGVLTKGKRALTPPVSDLKYSWKTWGKAKIFTPEARNSTFLIDGPDT
                     SECPNERRAWNSLEVEDYGFGMFTTNIWMKFREGSSEVCDHRLMSAAIKDQKAVHADM
                     GYWIESSKNQTWQIEKASLIEVKTCLWPKTHTLWSNGVLESQMLIPKSYAGPFSQHNY
                     RQGYATQTVGPWHLGKLEIDFGECPGTTVTIQEDCDHRGPSLRTTTASGKLVTQWCCR
                     SCTMPPLRFLGEDGCWYGMEIRPLSEKEENMVKSQVTAGQGTSETFSMGLLCLTLFVE
                     ECLRRRVTRKHMILVVVITLCAIILGGLTWMDLLRALIMLGDTMSGRIGGQIHLAIMA
                     VFKMSPGYVLGVFLRKLTSRETALMVIGMAMTTVLSIPHDLMELIDGISLGLILLKIV
                     TQFDNTQVGTLALSLTFIRSTMPLVMAWRTIMAVLFVVTLIPLCRTSCLQKQSHWVEI
                     TALILGAQALPVYLMTLMKGASRRSWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMVA
                     GGLLLAAYVMSGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEET
                     NMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQRSGALWDVPSPAATKKAALSE
                     GVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMIS
                     YGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSG
                     SPIINRKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVDEDIFRKKRLTIMDLH
                     PGAGKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEHT
                     GREIVDLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEAA
                     AIFMTATPPGATDPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAG
                     NDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPR
                     RCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDED
                     HAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGDL
                     PVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARV
                     YADPMALKDFKEFASGRKSITLDILTEIASLPTYLSSRAKLALDNIVMLHTTERGGRA
                     YQHALNELPESLETLMLVALLGAMTAGIFLFFMQGKGIGKLSMGLITIAVASGLLWVA
                     EIQPQWIAASIILEFFLMVLLIPEPEKQRTPQDNQLIYVILTLITIIGLIAANEMGLI
                     EKTKTDFGFYQVKTETTILDVDLRPASAWTLYAVATTILTPMLRHTIENTSANLSLAA
                     IANQAAVLMGLGKGWPLHRMDLGVPLLAMGCYSQVNPTTLTASLVMLLVHYAIIGPGL
                     QAKATREAQKRTAAGIMKNPTVDGITVIDLEPISYDPKFEKQLGQVMLLVLCAGQLLL
                     MRTTWAFCEVLTLATGPILTLWEGNPGRFWNTTIAVSTANIFRGSYLAGAGLAFSLIK
                     NAQTPRRGTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSKI
                     KHAVSRGSSKIRWIVERGMVKPKGKVVDLGCGRGGWSYYMATLKNVTEVKGYTKGGPG
                     HEEPIPMATYGWNLVKLHSGVDVFYKPTEQVDTLLCDIGESSSNPTIEEGRTLRVLKM
                     VEPWLSSKPEFCIKVLNPYMPTVIEELEKLQRKHGGNLVRCPLSRNSTHEMYWVSGAS
                     GNIVSSVNTTSKMLLNRFTTRHRKPTYEKDVDLGAGTRSVSTETEKPDMTIIGRRLQR
                     LQEEHKETWHYDQENPYRTWAYHGSYEAPSTGSASSMVNGVVKLLTKPWDVIPMVTQL
                     AMTDTTPFGQQRVFKEKVDTRTPQPKPGTRMVMTTTANWLWALLGKKKNPRLCTREEF
                     ISKVRSNAAIGAVFQEEQGWTSASEAVNDSRFWELVDKERALHQEGKCESCVYNMMGK
                     REKKLGEFGRAKGSRAIWYMWLGARFLEFEALGFLNEDHWFGRENSWSGVEGEGLHRL
                     GYILEEIDKKDGDLMYADDTAGWDTRITEDDLQNEELITEQMAPHHKILAKAIFKLTY
                     QNKVVKVLRPTPRGAVMDIISRKDQRGSGQVGTYGLNTFTNMEVQLIRQMEAEGVITQ
                     DDMQNPKGLKERVEKWLKECGVDRLKRMAISGDDCVVKPLDERFGTSLLFLNDMGKVR
                     KDIPQWEPSKGWKNWQEVPFCSHHFHKIFMKDGRSLVVPCRNQDELIGRARISQGAGW
                     SLRETACLGKAYAQMWSLMYFHRRDLRLASMAICSAVPTEWFPTSRTTWSIHAHHQWM
                     TTEDMLKVWNRVWIEDNPNMTDKTPVHSWEDIPYLGKREDLWCGSLIGLSSRATWAKN
                     IHTAITQVRNLIGKEEYVDYMPVMKRYSAPSESEGVL"
mat_peptide          102. .440
                     /product="anchored capsid protein"
                     /note="anchC"
mat_peptide          102. .398
                     /product="virion capsid protein"
                     /note="virC"
mat_peptide          441. .938
                     /product="membrane precursor protein"
                     /note="prM"
mat_peptide          714. .938
                     /product="membrane protein"
                     /note="M"
mat_peptide          939. .2423
                     /product="envelope protein"
                     /note="E"
mat_peptide          2424. .3479
                     /product="NS1 protein"
mat_peptide          3480. .4133
                     /product="NS2A protein"
mat_peptide          4134. .4523
                     /product="NS2B protein"
mat_peptide          4524. .6377
                     /product="NS3 protein"
mat_peptide          6378. .6758
                     /product="NS4A protein"
mat_peptide          6759. .6827
                     /product="2K protein"
mat_peptide          6828. .7562
                     /product="NS4B protein"
mat_peptide          7563. .10262
                     /product="NS5 protein"
misc_feature         10619. .10627
                     /note="KpnI/AgeI linker"
promoter             14990. .15239
                     /note="SP6 promoter"
ORIGIN

  Query Match            34.0%;  Score 1636.8;  DB 11;  Length 15239;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    | |     |  ||||||||||||||||||||||||||||||
Db      11157 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11216

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11217 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11276

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11277 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11336

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11337 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11396

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11397 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11456

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11457 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11516

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11517 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11576

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      11577 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11636

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11637 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11696

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11697 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11756

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11757 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11816
Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11817 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11876
Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11877 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11936
Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11937 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 11996
Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11997 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12056
Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12057 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12116
Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12117 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12176
Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12177 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12236
Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12237 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12296
Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12297 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12356
Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12357 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12416
Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12417 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12476
Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      12477 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12536
Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12537 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12596
Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12597 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12656
Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12657 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12716
Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12717 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12776
Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12777 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12836
Qy       2777 ------------------------------------------------------------ 2776
Db      12837 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12896
```

```
Qy       2777 ------------------------------------------------------------ 2776
Db      12897 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12956
Qy       2777 ------------------------------------------------------------ 2776
Db      12957 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13016
Qy       2777 ------------------------------------------------------------ 2776
Db      13017 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13076
Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||||
Db      13077 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13136
Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13137 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13196
Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13197 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13256
Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13257 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13316
Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13317 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13376
Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
             |||||||||||||||||||||||||  ||| 
Db      13377 CTGGACGCGGATGAACAGGCAGACATCT 13404
```

```
RESULT 9
AY656168
LOCUS       AY656168               15256 bp    DNA     circular SYN 14-JAN-2005
DEFINITION  Chimeric dengue virus vector p4(delta30)-D3L-ME, complete sequence.
ACCESSION   AY656168
VERSION     AY656168.1  GI:50345910
KEYWORDS    .
SOURCE      Chimeric dengue virus vector p4(delta30)-D3L-ME
  ORGANISM  Chimeric dengue virus vector p4(delta30)-D3L-ME
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15256)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Firestone,C.Y., Hanley,K.A.,
            Murphy,B.R. and Whitehead,S.S.
  TITLE     Genetically modified, live attenuated dengue virus type 3 vaccine
            candidates
  JOURNAL   Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
   PUBMED   15642976
REFERENCE   2  (bases 1 to 15256)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15256
                     /organism="Chimeric dengue virus vector
                     p4(delta30)-D3L-ME"
                     /mol_type="other DNA"
                     /db_xref="taxon:284998"
                     /focus
     source          1. .401
                     /organism="Dengue virus type 4"
                     /mol_type="other DNA"
                     /db_xref="taxon:11070"
     source          402. .2339
                     /organism="Dengue virus type 3"
```

```
                     /mol_type="other DNA"
                     /db_xref="taxon:11069"
                     /note="Sleman/78"
     source          2363. .10635
                     /organism="Dengue virus type 4"
                     /mol_type="other DNA"
                     /db_xref="taxon:11070"
     source          10645. .15006
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     source          15007. .15256
                     /organism="Cloning vector pGEM-3"
                     /mol_type="other DNA"
                     /db_xref="taxon:90108"
     misc_feature    102. .10279
                     /note="encodes polyprotein precursor once XhoI linker
                     sequence is removed prior to preparation of infectious
                     transcripts corresponding to the viral genome"
     misc_feature    2340. .2362
                     /note="XhoI linker; required for stable replication in
                     Escherichia coli"
     misc_feature    10636. .10644
                     /note="KpnI/AgeI linker"
     promoter        15238. .15256
                     /note="SP6"
ORIGIN

  Query Match             34.0%;  Score 1636.8;  DB 11;  Length 15256;
  Best Local Similarity   86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    |  |     |  ||||||||||||||||||||||||||||||
Db      11174 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11233

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11234 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11293

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11294 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11353

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11354 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11413

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11414 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11473

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11474 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11533

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11534 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11593

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      11594 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11653

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11654 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11713

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11714 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11773

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11774 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11833

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11834 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11893

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11894 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11953

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11954 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12013

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12014 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12073

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12074 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12133

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12134 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12193

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12194 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12253

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12254 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12313

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12314 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12373

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12374 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12433

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12434 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12493

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12494 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12553

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12554 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12613

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12614 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12673

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12674 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12733

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12734 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12793

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12794 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12853

Qy       2777 ---------------------------------------------------------- 2776
```

```
Db      12854 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12913

Qy       2777 ------------------------------------------------------------ 2776

Db      12914 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12973

Qy       2777 ------------------------------------------------------------ 2776

Db      12974 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13033

Qy       2777 ------------------------------------------------------------ 2776

Db      13034 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13093

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||
Db      13094 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13153

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13154 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13213

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13214 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13273

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13274 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13333

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13334 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13393

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
              |||||||||||||||||||||||| |||
Db      13394 CTGGACGCGGATGAACAGGCAGACATCT 13421
```

```
RESULT 10
AY243468
LOCUS       AY243468               15268 bp    DNA     circular SYN 20-OCT-2003
DEFINITION  Chimeric Dengue virus vector p4-D2-CME, complete sequence.
ACCESSION   AY243468
VERSION     AY243468.1  GI:30026603
KEYWORDS    .
SOURCE      Chimeric Dengue virus vector p4-D2-CME
  ORGANISM  Chimeric Dengue virus vector p4-D2-CME
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15268)
  AUTHORS   Whitehead,S.S., Hanley,K.A., Blaney,J.E. Jr., Gilmore,L.E.,
            Elkins,W.R. and Murphy,B.R.
  TITLE     Substitution of the structural genes of dengue virus type 4 with
            those of type 2 results in chimeric vaccine candidates which are
            attenuated for mosquitoes, mice, and rhesus monkeys
  JOURNAL   Vaccine 21 (27-28), 4307-4316 (2003)
   PUBMED   14505913
REFERENCE   2  (bases 1 to 15268)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2003) LID, NIAID, 50 South Dr., Room 6515,
            Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15268
                     /organism="Chimeric Dengue virus vector p4-D2-CME"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:227157"
                     /focus
     source          1. .88
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     source          89. .2343
                     /organism="Dengue virus type 2 (NGC-prototype)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11065"
     source          2344. .10647
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     CDS             97. .10263
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAP06254.1"
                     /db_xref="GI:30026604"
                     /translation="MNNQRKKARNTPFNMLKRERNRVSTVQQLTKRFSLGMLQGRGPL
                     KLFMALVAFLRFLTIPPTAGILKRWGTIKKSKAINVLRGFRKEIGRMLNILNRRRRTA
                     GMIIMLIPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTEDGVNMCTLMAMDLGELCE
                     DTITYKCPLLRQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLE
                     TRTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRALIFILLTAVA
                     PSMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPA
                     TLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGG
                     IVTCAMFTCKKNMEGKVVQPENLEYTIVITPHSGEEHAVGNDTGKHGKEIKITPQSSI
                     TEAELTGYGTVTMECSPRTGLDFNEMVLLQMENKAWLVHRQWFLDLPLPWLPGADTQG
                     SNWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCR
                     LRMDKLQLKGMSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEK
                     RHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMF
                     ETTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIG
                     VIITWIGMNSRNTSMAMTCIAVGGITLFLGFTVQADMGCVASWSGKELKCGSGIFVVD
                     NVHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLENVMWKQITNELNYVLWE
                     GGHDLTVVAGDVKGVLTKGKRALTPPVSDLKYSWKTWGKAKIFTPEARNSTFLIDGPD
                     TSECPNERRAWNSLEVEDYGFGMFTTNIWMKFREGSSEVCDHRLMSAAIKDQKAVHAD
                     MGYWIESSKNQTWQIEKASLIEVKTCLWPKTHTLWSNGVLESQMLIPKSYAGPFSQHN
                     YRQGYATQTVGPWHLGKLEIDFGECPGTTVTIQEDCDHRGPSLRTTTASGKLVTQWCC
                     RSCTMPPLRFLGEDGCWYGMEIRPLSEKEENMVKSQVTAGQGTSETFSMGLLCLTLFV
                     EECLRRRVTRKHMILVVVITLCAIILGGLTWMDLLRALIMLGDTMSGRIGGQIHLAIM
                     AVFKMSPGYVLGVFLRKLTSRETALMVIGMAMTTVLSIPHDLMELIDGISLGLILLKI
                     VTQFDNTQVGTLALSLTFIRSTMPLVMAWRTIMAVLFVVTLIPLCRTSCLQKQSHWVE
                     ITALILGAQALPVYLMTLMKGASRRSWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMV
                     AGGLLLAAYVMSGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEE
                     TNMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQRSGALWDVPSPAATKKAALS
                     EGVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMI
                     SYGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTS
                     GSPIINRKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVDEDIFRKKRLTIMDL
                     HPGAGKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEH
                     TGREIVDLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEA
                     AAIFMTATPPGATDPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKA
                     GNDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDP
                     RRCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDE
                     DHAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGD
                     LPVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDAR
                     VYADPMALKDFKEFASGRKSITLDILTEIASLPTYLSSRAKLALDNIVMLHTTERGGR
                     AYQHALNELPESLETLMLVALLGAMTAGIFLFFMQGKGIGKLSMGLITIAVASGLLWV
                     AEIQPQWIAASIILEFFLMVLLIPEPEKQRTPQDNQLIYVILTILTIIGLIAANEMGL
                     IEKTKTDFGFYQVKTETTILDVDLRPASAWTLYAVATTILTPMLRHTIENTSANLSLA
                     AIANQAAVLMGLGKGWPLHRMDLGVPLLAMGCYSQVNPTTLTASLVMLLVHYAIIGPG
                     LQAKATREAQKRTAAGIMKNPTVDGITVIDLEPISYDPKFEKQLGQVMLLVLCAGQLL
                     LMRTTWAFCEVLTLATGPILTLWEGNPGRFWNTTIAVSTANIFRGSYLAGAGLAFSLI
                     KNAQTPRRGTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSK
                     IKHAVSRGSSKIRWIVERGMVKPKGKVVDLGCGRGGWSYYMATLKNVTEVKGYTKGGP
                     GHEEPIPMATYGWNLVKLHSGVDVFYKPTEQVDTLLCDIGESSSNPTIEEGRTLRVLK
                     MVEPWLSSKPEFCIKVLNPYMPTVIEELEKLQRKHGGNLVRCPLSRNSTHEMYWVSGA
                     SGNIVSSVNTTSKMLLNRFTTRHRKPTYEKDVDLGAGTRSVSTETEKPDMTIIGRRLQ
                     RLQEEHKETWHYDQENPYRTWAYHGSYEAPSTGSASSMVNGVVKLLTKPWDVIPMVTQ
                     LAMTDTTPFGQQRVFKEKVDTRTPQPKPGTRMVMTTTANWLWALLGKKKNPRLCTREE
                     FISKVRSNAAIGAVFQEEQGWTSASEAVNDSRFWELVDKERALHQEGKCESCVYNMMG
                     KREKKLGEFGRAKGSRAIWYMWLGARFLEFEALGFLNEDHWFGRENSWSGVEGEGLHR
                     LGYILEEIDKKDGDLMYADDTAGWDTRITEDDLQNEELITEQMAPHHKILAKAIFKLT
                     YQNKVVKVLRPTPRGAVMDIISRKDQRGSGQVGTYGLNTFTNMEVQLIRQMEAEGVIT
                     QDDMQNPKGLKERVEKWLKECGVDRLKRMAISGDDCVVKPLDERFGTSLLFLNDMGKV
                     RKDIPQWEPSKGWKNWQEVPFCSHHFHKIFMKDGRSLVVPCRNQDELIGRARISQGAG
                     WSLRETACLGKAYAQMWSLMYFHRRDLRLASMAICSAVPTEWFPTSRTTWSIHAHHQW
                     MTTEDMLKVWNRVWIEDNPNMTDKTPVHSWEDIPYLGKREDLWCGSLIGLSSRATWAK
                     NIHTAITQVRNLIGKEEYVDYMPVMKRYSAPSESEGVL"
     mat_peptide     97. .438
                     /product="anchored capsid protein"
```

```
                     /note="anchC"
     mat_peptide     97. .396
                     /product="virion capsid protein"
                     /note="virC"
     mat_peptide     439. .936
                     /product="membrane precursor protein"
                     /note="prM"
     mat_peptide     712. .936
                     /product="membrane protein"
                     /note="M"
     mat_peptide     937. .2421
                     /product="envelope protein"
                     /note="E"
     mat_peptide     2422. .3477
                     /product="NS1 protein"
     mat_peptide     3478. .4131
                     /product="NS2A protein"
     mat_peptide     4132. .4521
                     /product="NS2B protein"
     mat_peptide     4522. .6375
                     /product="NS3 protein"
     mat_peptide     6376. .6756
                     /product="NS4A protein"
     mat_peptide     6757. .6825
                     /product="2K protein"
     mat_peptide     6826. .7560
                     /product="NS4B protein"
     mat_peptide     7561. .10260
                     /product="NS5 protein"
ORIGIN

  Query Match           34.0%;  Score 1636.8;  DB 11;  Length 15268;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |     |  |       |  ||||||||||||||||||||||||||||
Db      11186 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11245

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11246 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11305

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11306 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11365

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11366 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11425

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11426 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11485

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11486 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11545

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11546 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11605

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      11606 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11665

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11666 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11725

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      11726 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11785

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11786 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11845

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11846 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11905

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11906 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11965

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11966 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12025

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12026 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12085

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12086 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12145

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12146 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12205

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12206 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12265

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12266 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12325

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12326 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12385

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12386 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12445

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12446 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12505

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      12506 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12565

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12566 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12625

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12626 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12685

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12686 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12745

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12746 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12805

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12806 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12865
```

```
Qy       2777 ------------------------------------------------------------ 2776

Db      12866 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12925

Qy       2777 ------------------------------------------------------------ 2776

Db      12926 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12985

Qy       2777 ------------------------------------------------------------ 2776

Db      12986 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13045

Qy       2777 ------------------------------------------------------------ 2776

Db      13046 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13105

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||||
Db      13106 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13165

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13166 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13225

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13226 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13285

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13286 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13345

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13346 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13405

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
              |||||||||||||||||||||||| |||
Db      13406 CTGGACGCGGATGAACAGGCAGACATCT 13433
```

```
RESULT 11
AY243466
LOCUS       AY243466               15270 bp    DNA     circular SYN 20-OCT-2003
DEFINITION  Chimeric Dengue virus vector p4-D2-ME, complete sequence.
ACCESSION   AY243466
VERSION     AY243466.1  GI:30026599
KEYWORDS    .
SOURCE      Chimeric Dengue virus vector p4-D2-ME
  ORGANISM  Chimeric Dengue virus vector p4-D2-ME
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15270)
  AUTHORS   Whitehead,S.S., Hanley,K.A., Blaney,J.E. Jr., Gilmore,L.E.,
            Elkins,W.R. and Murphy,B.R.
  TITLE     Substitution of the structural genes of dengue virus type 4 with
            those of type 2 results in chimeric vaccine candidates which are
            attenuated for mosquitoes, mice, and rhesus monkeys
  JOURNAL   Vaccine 21 (27-28), 4307-4316 (2003)
   PUBMED   14505913
REFERENCE   2  (bases 1 to 15270)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2003) LID, NIAID, 50 South Dr., Room 6515,
            Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15270
                     /organism="Chimeric Dengue virus vector p4-D2-ME"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:227155"
                     /focus
     source          1. .404
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     source          405. .2345
                     /organism="Dengue virus type 2 (NGC-prototype)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11065"
     source          2346. .10649
                     /organism="Dengue virus type 4"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:11070"
     CDS             102. .10265
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAP06252.1"
                     /db_xref="GI:30026600"
                     /translation="MNQRKKVVRPPFNMLKRERNRVSTPQGLVKRFSTGLFSGKGPLR
                     MVLAFITFLRVLSIPPTAGILKRWGQLKKNKAIKILIGFRKEIGRMLNILNGRKRSAG
                     MIIMLIPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTEDGVNMCTLMAMDLGELCED
                     TITYKCPLLRQNEPEDIDCWCNSTSTWVTYGTCTTTGEHRREKRSVALVPHVGMGLET
                     RTETWMSSEGAWKHAQRIETWILRHPGFTIMAAILAYTIGTTHFQRALIFILLTAVAP
                     SMTMRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPAT
                     LRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGGI
                     VTCAMFTCKKNMEGKVVQPENLEYTIVITPHSGEEHAVGNDTGKHGKEIKITPQSSIT
                     EAELTGYGTVTMECSPRTGLDFNEMVLLQMENKAWLVHRQWFLDLPLPWLPGADTQGS
                     NWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCRL
                     RMDKLQLKGMSYSMCTGKFKVVKEIAETQHGTIVIRVQYEGDGSPCKIPFEIMDLEKR
                     HVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFE
                     TTMRGAKRMAILGDTAWDFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIGV
                     IITWIGMNSRNTSMAMTCIAVGGITLFLGFTVQADMGCVASWSGKELKCGSGIFVVDN
                     VHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLENVMWKQITNELNYVLWEG
                     GHDLTVVAGDVKGVLTKGKRALTPPVSDLKYSWKTWGKAKIFTPEARNSTFLIDGPDT
                     SECPNERRAWNSLEVEDYGFGMFTTNIWMKFREGSSEVCDHRLMSAAIKDQKAVHADM
                     GYWIESSKNQTWQIEKASLIEVKTCLWPKTHTLWSNGVLESQMLIPKSYAGPFSQHNY
                     RQGYATQTVGPWHLGKLEIDFGECPGTTVTIQEDCDHRGPSLRTTTASGKLVTQWCCR
                     SCTMPPLRFLGEDGCWYGMEIRPLSEKEENMVKSQVTAGQGTSETFSMGLLCLTLFVE
                     ECLRRRVTRKHMILVVVITLCAIILGGLTWMDLLRALIMLGDTMSGRIGGQIHLAIMA
                     VFKMSPGYVLGVFLRKLTSRETALMVIGMAMTTVLSIPHDLMELIDGISLGLILLKIV
                     TQFDNTQVGTLALSLTFIRSTMPLVMAWRTIMAVLFVVTLIPLCRTSCLQKQSHWVEI
                     TALILGAQALPVYLMTLMKGASRRSWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMVA
                     GGLLLAAYVMSGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEET
                     NMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQRSGALWDVPSPAATKKAALSE
                     GVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMIS
                     YGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSG
                     SPIINRKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVDEDIFRKKRLTIMDLH
                     PGAGKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEHT
                     GREIVDLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEAA
                     AIFMTATPPGATDPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAG
                     NDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPR
                     RCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDED
                     HAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGDL
                     PVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARV
                     YADPMALKDFKEFASGRKSITLDILTEIASLPTYLSSRAKLALDNIVMLHTTERGGRA
                     YQHALNELPESLETLMLVALLGAMTAGIFLFFMQGKGIGKLSMGLITIAVASGLLWVA
                     EIQPQWIAASIILEFFLMVLLIPEPEKQRTPQDNQLIYVILTILTIIGLIAANEMGLI
                     EKTKTDFGFYQVKTETTILDVDLRPASAWTLYAVATTILTPMLRHTIENTSANLSLAA
                     IANQAAVLMGLGKGWPLHRMDLGVPLLAMGCYSQVNPTTLTASLVMLLVHYAIIGPGL
                     QAKATREAQKRTAAGIMKNPTVDGITVIDLEPISYDPKFEKQLGQVMLLVLCAGQLLL
                     MRTTWAFCEVLTLATGPILTLWEGNPGRFWNTTIAVSTANIFRGSYLAGAGLAFSLIK
                     NAQTPRRGTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSKI
                     KHAVSRGSSKIRWIVERGMVKPKGKVVDLGCGRGGWSYYMATLKNVTEVKGYTKGGPG
                     HEEPIPMATYGWNLVKLHSGVDVFYKPTEQVDTLLCDIGESSSNPTIEEGRTLRVLKM
                     VEPWLSSKPEFCIKVLNPYMPTVIEELEKLQRKHGGNLVRCPLSRNSTHEMYWVSGAS
                     GNIVSSVNTTSKMLLNRFTTRHRKPTYEKDVDLGAGTRSVSTETEKPDMTIIGRRLQR
                     LQEEHKETWHYDQENPYRTWAYHGSYEAPSTGSASSMVNGVVKLLTKPWDVIPMVTQL
                     AMTDTTPFGQQRVFKEKVDTRTPQPKPGTRMVMTTTANWLWALLGKKKNPRLCTREEF
                     ISKVRSNAAIGAVFQEEQGWTSASEAVNDSRFWELVDKERALHQEGKCESCVYNMMGK
                     REKKLGEFGRAKGSRAIWYMWLGARFLEFEALGFLNEDHWFGRENSWSGVEGEGLHRL
                     GYILEEIDKKDGDLMYADDTAGWDTRITEDDLQNEELITEQMAPHHKILAKAIFKLTY
                     QNKVVKVLRPTPRGAVMDIISRKDQRGSGQVGTYGLNTFTNMEVQLIRQMEAEGVITQ
                     DDMQNPKGLKERVEKWLKECGVDRLKRMAISGDDCVVKPLDERFGTSLLFLNDMGKVR
                     KDIPQWEPSKGWKNWQEVPFCSHHFHKIFMKDGRSLVVPCRNQDELIGRARISQGAGW
                     SLRETACLGKAYAQMWSLMYFHRRDLRLASMAICSAVPTEWFPTSRTTWSIHAHHQWM
                     TTEDMLKVWNRVWIEDNPNMTDKTPVHSWEDIPYLGKREDLWCGSLIGLSSRATWAKN
```

```
                     IHTAITQVRNLIGKEEYVDYMPVMKRYSAPSESEGVL"
     mat_peptide     102. .440
                     /product="anchored capsid protein"
                     /note="anchC"
     mat_peptide     102. .398
                     /product="virion capsid protein"
                     /note="virC"
     mat_peptide     441. .938
                     /product="membrane precursor protein"
                     /note="prM"
     mat_peptide     714. .938
                     /product="membrane protein"
                     /note="M"
     mat_peptide     939. .2423
                     /product="envelope protein"
                     /note="E"
     mat_peptide     2424. .3479
                     /product="NS1 protein"
     mat_peptide     3480. .4133
                     /product="NS2A protein"
     mat_peptide     4134. .4523
                     /product="NS2B protein"
     mat_peptide     4524. .6377
                     /product="NS3 protein"
     mat_peptide     6378. .6758
                     /product="NS4A protein"
     mat_peptide     6759. .6827
                     /product="2K protein"
     mat_peptide     6828. .7562
                     /product="NS4B protein"
     mat_peptide     7563. .10262
                     /product="NS5 protein"
ORIGIN

  Query Match           34.0%;  Score 1636.8;  DB 11;  Length 15270;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
Db      11188 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11247

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
Db      11248 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11307

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
Db      11308 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11367

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
Db      11368 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11427

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
Db      11428 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11487

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
Db      11488 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11547

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
Db      11548 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11607

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
Db      11608 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11667

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
Db      11668 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11727

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
Db      11728 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11787

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
Db      11788 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11847

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
Db      11848 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11907

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
Db      11908 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11967

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
Db      11968 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12027

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
Db      12028 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12087

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
Db      12088 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12147

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
Db      12148 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12207

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
Db      12208 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12267

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
Db      12268 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12327

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
Db      12328 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12387

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
Db      12388 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12447

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
Db      12448 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12507

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
Db      12508 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12567

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
Db      12568 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12627

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
Db      12628 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12687

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
Db      12688 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12747

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
Db      12748 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12807
```

```
Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG---------- 2776
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12808 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12867

Qy       2777 ------------------------------------------------------------ 2776

Db      12868 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12927

Qy       2777 ------------------------------------------------------------ 2776

Db      12928 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12987

Qy       2777 ------------------------------------------------------------ 2776

Db      12988 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13047

Qy       2777 ------------------------------------------------------------ 2776

Db      13048 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13107

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||
Db      13108 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13167

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13168 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13227

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13228 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13287

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13288 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13347

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13348 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13407

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
             ||||||||||||||||||||||   |||
Db      13408 CTGGACGCGGATGAACAGGCAGACATCT 13435
```

```
RESULT 12
AY648301
LOCUS       AY648301               15270 bp    DNA     circular SYN 14-JAN-2005
DEFINITION  Dengue virus type 4 vector p4, complete sequence.
ACCESSION   AY648301
VERSION     AY648301.1  GI:49781322
KEYWORDS    .
SOURCE      Dengue virus type 4 vector p4
  ORGANISM  Dengue virus type 4 vector p4
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15270)
  AUTHORS   Durbin,A.P., Karron,R.A., Sun,W., Vaughn,D.W., Reynolds,M.J.,
            Perreault,J.R., Thumar,B., Men,R., Lai,C.J., Elkins,W.R.,
            Chanock,R.M., Murphy,B.R. and Whitehead,S.S.
  TITLE     Attenuation and immunogenicity in humans of a live dengue virus
            type-4 vaccine candidate with a 30 nucleotide deletion in its
            3'-untranslated region
  JOURNAL   Am. J. Trop. Med. Hyg. 65 (5), 405-413 (2001)
   PUBMED   11716091
REFERENCE   2  (bases 1 to 15270)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Firestone,C.Y., Hanley,K.A.,
            Murphy,B.R. and Whitehead,S.S.
  TITLE     Genetically modified, live attenuated dengue virus type 3 vaccine
            candidates
  JOURNAL   Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
   PUBMED   15642976
REFERENCE   3  (bases 1 to 15270)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15270
                     /organism="Dengue virus type 4 vector p4"
                     /mol_type="other DNA"
                     /db_xref="taxon:283795"
                     /focus
     source          1. .10649
                     /organism="Dengue virus type 4"
                     /mol_type="other DNA"
                     /db_xref="taxon:11070"
                     /country="Dominica"
     source          10659. .15020
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     source          15021. .15270
                     /organism="Cloning vector pGEM-3"
                     /mol_type="other DNA"
                     /db_xref="taxon:90108"
     CDS             102. .10265
                     /codon_start=1
                     /product="polyprotein precursor"
                     /protein_id="AAT68472.1"
                     /db_xref="GI:49781323"
                     /translation="MNQRKKVVRPPFNMLKRERNRVSTPQGLVKRFSTGLFSGKGPLR
                     MVLAFITFLRVLSIPPTAGILKRWGQLKKNKAIKLIGFRKEIGRMLNILNGRKRSTI
                     TLLCLIPTVMAFSLSTRDGEPLMIVAKHERGRPLLFKTTEGINKCTLIAMDLGEMCED
                     TVTYKCPLLVNTEPEDIDCWCNLTSTWVMYGTCTQSGERRREKRSVALTPHSGMGLET
                     RAETWMSSEGAWKHAQRVESWILRNPGFALLAGFMAYMIGQTGIQRTVFFVLMMLVAP
                     SYGMRCVGVGNRDFVEGVSGGAWVDLVLEHGGCVTTMAQGKPTLDFELTKTTAKEVAL
                     LRTYCIEASISNITTATRCPTQGEPYLKEEQDQQYICRRDVVDRGWGNGCGLFGKGGV
                     VTCAKFSCSGKITGNLVQIENLEYTVVVTVHNGDTHAVGNDTSNHGVTAMITPRSPSV
                     EVKLPDYGELTLDCEPRSGIDFNEMILMKMKKKTWLVHKQWFLDLPLPWTAGADTSEV
                     HWNYKERMVTFKVPHAKRQDVTVLGSQEGAMHSALAGATEVDSGDGNHMFAGHLKCKV
                     RMEKLRIKGMSYTMCSGKFSIDKEMAETQHGTTVVKVKYEGAGAPCKVPIEIRDVNKE
                     KVVGRIISSTPLAENTNSVTNIELEPPFGDSYIVIGVGNSALTLHWFRKGSSIGKMFE
                     STYRGAKRMAILGETAWDFGSVGGLFTSLGKAVHQVFGSVYTTMFGGVSWMIRILIGF
                     LVLWIGTNSRNTSMAMTCIAVGGITLFLGFTVQADMGCVASWSGKELKCGSGIFVVDN
                     VHTWTEQYKFQPESPARLASAILNAHKDGVCGIRSTTRLENVMWKQITNELNYVLWEG
                     GHDLTVVAGDVKGVLTKGKRALTPPVSDLKYSWKTWGKAKIFTPEARNSTFLIDGPDT
                     SECPNERRAWNSLEVEDYGFGMFTTNIWMKFREGSSEVCDHRLMSAAIKDQKAVHADM
                     GYWIESSKNQTWQIEKASLIEVKTCLWPKTHTLWSNGVLESQMLIPKSYAGPFSQHNY
                     RQGYATQTVGPWHLGKLEIDFGECPGTTVTIQEDCDHRGPSLRTTTASGKLVTQWCCR
                     SCTMPPLRFLGEDGCWYGMEIRPLSEKEENMVKSQVTAGQGTSETFSMGLLCLTLFVE
                     ECLRRRVTRKHMILVVVITLCAIILGGLTWMDLLRALIMLGDTMSGRIGGQIHLAIMA
                     VFKMSPGYVLGVFLRKLTSRETALMVIGMAMTTVLSIPHDLMELIDGISLGLILLKIV
                     TQFDNTQVGTLALSLTFIRSTMPLVMAWRTIMAVLFVVTLIPLCRTSCLQKQSHWVEI
                     TALILGAQALPVYLMTLMKGASRRSWPLNEGIMAVGLVSLLGSALLKNDVPLAGPMVA
                     GGLLLAAYVMSGSSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEET
                     NMITLLVKLALITVSGLYPLAIPVTMTLWYMWQVKTQRSGALWDVPSPAATKKAALSE
                     GVYRIMQRGLFGKTQVGVGIHMEGVFHTMWHVTRGSVICHETGRLEPSWADVRNDMIS
                     YGGGWRLGDKWDKEEDVQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSG
                     SPIINRKGKVIGLYGNGVVTKSGDYVSAITQAERIGEPDYEVDEDIFRKKRLTIMDLH
                     PGAGKTKRILPSIVREALKRRLRTLILAPTRVVAAEMEEALRGLPIRYQTPAVKSEHT
                     GREIVDLMCHATFTTRLLSSTRVPNYNLIVMDEAHFTDPSSVAARGYISTRVEMGEAA
                     AIFMTATPPGATDPFPQSNSPIEDIEREIPERSWNTGFDWITDYQGKTVWFVPSIKAG
                     NDIANCLRKSGKKVIQLSRKTFDTEYPKTKLTDWDFVVTTDISEMGANFRAGRVIDPR
                     RCLKPVILPDGPERVILAGPIPVTPASAAQRRGRIGRNPAQEDDQYVFSGDPLKNDED
                     HAHWTEAKMLLDNIYTPEGIIPTLFGPEREKTQAIDGEFRLRGEQRKTFVELMRRGDL
                     PVWLSYKVASAGISYKDREWCFTGERNNQILEENMEVEIWTREGEKKKLRPRWLDARV
                     YADPMALKDFKEFASGRKSITLDILTEIASLPTYLSSRAKLALDNIVMLHTTERGGRA
                     YQHALNELPESLETLMLVALLGAMTAGIFLFFMQGKGIGKLSMGLITIAVASGLLWVA
                     EIQPQWIAASIILEFFLMVLLIPEPEKQRTPQDNQLIYVILTILTIIGLIAANEMGLI
                     EKTKTDFGFYQVKTETTILDVDLRPASAWTLYAVATTILTPMLRHTIENTSANLSLAA
                     IANQAAVLMGLGKGWPLHRMDLGVPLLAMGCYSQVNPTTLTASLVMLLVHYAIIGPGL
                     QAKATREAQKRTAAGIMKNPTVDGITVIDLEPISYDPKFEKQLGQVMLLVLCAGQLLL
                     MRTTWAFCEVLTLATGPILTLWEGNPGRFWNTTIAVSTANIFRGSYLAGAGLAFSLIK
                     NAQTPRRGTGTTGETLGEKWKRQLNSLDRKEFEEYKRSGILEVDRTEAKSALKDGSKI
                     KHAVSRGSSKIRWIVERGMVKPKGKVVDLGCGRGGWSYYMATLKNVTEVKGYTKGGPG
                     HEEPIPMATYGWNLVKLHSGVDVFYKPTEQVDTLLCDIGESSSNPTIEEGRTLRVLKM
```

```
                     VEPWLSSKPEFCIKVLNPYMPTVIEELEKLQRKHGGNLVRCPLSRNSTHEMYWVSGAS
                     GNIVSSVNTTSKMLLNRFTTRHRKPTYEKDVDLGAGTRSVSTETEKPDMTIIGRRLQR
                     LQEEHKETWHYDQENPYRTWAYHGSYEAPSTGSASSMVNGVVKLLTKPWDVIPMVTQL
                     AMTDTTPFGQQRVFKEKVDTRTPQPKPGTRMVMTTTANWLWALLGKKKNPRLCTREEF
                     ISKVRSNAAIGAVFQEEQGWTSASEAVNDSRFWELVDKERALHQEGKCESCVYNMMGK
                     REKKLGEFGRAKGSRAIWYMWLGARFLEFEALGFLNEDHWFGRENSWSGVEGEGLHRL
                     GYILEEIDKKDGDLMYADDTAGWDTRITEDDLQNEELITEQMAPHHKILAKAIFKLTY
                     QNKVVKVLRPTPRGAVMDIISRKDQRGSGQVGTYGLNTFTNMEVQLIRQMEAEGVITQ
                     DDMQNPKGLKERVEKWLKECGVDRLKRMAISGDDCVVKPLDERFGTSLLFLNDMGKVR
                     KDIPQWEPSKGWKNWQEVPFCSHHFHKIFMKDGRSLVVPCRNQDELIGRARISQGAGW
                     SLRETACLGKAYAQMWSLMYFHRRDLRLASMAICSAVPTEWFPTSRTTWSIHAHHQWM
                     TTEDMLKVWNRVWIEDNPNMTDKTPVHSWEDIPYLGKREDLWCGSLIGLSSRATWAKN
                     IHTAITQVRNLIGKEEYVDYMPVMKRYSAPSESEGVL"
     mat_peptide     102. .440
                     /product="anchored capsid protein"
                     /note="anchC"
     mat_peptide     102. .398
                     /product="virion capsid protein"
                     /note="virC"
     mat_peptide     441. .938
                     /product="membrane protein precursor"
                     /note="prM"
     mat_peptide     714. .938
                     /product="membrane protein"
                     /note="M"
     mat_peptide     939. .2423
                     /product="envelope protein"
                     /note="E"
     mat_peptide     2424. .3479
                     /product="NS1 protein"
     mat_peptide     3480. .4133
                     /product="NS2A protein"
     mat_peptide     4134. .4523
                     /product="NS2B protein"
     mat_peptide     4524. .6377
                     /product="NS3 protein"
     mat_peptide     6378. .6758
                     /product="NS4A protein"
     mat_peptide     6759. .6827
                     /product="2K protein"
     mat_peptide     6828. .7562
                     /product="NS4B protein"
     mat_peptide     7563. .10262
                     /product="NS5 protein"
     misc_feature    10650. .10658
                     /note="KpnI/AgeI linker"
     promoter        15021. .15270
                     /note="SP6"
ORIGIN

  Query Match           34.0%;  Score 1636.8;  DB 11;  Length 15270;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    |  |      | ||||||||||||||||||||||||||||||
Db      11188 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11247

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11248 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11307

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11308 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11367

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11368 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11427

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11428 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11487

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11488 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11547

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11548 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11607

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      11608 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11667

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11668 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11727

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11728 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11787

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11788 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11847

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11848 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11907

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11908 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 11967

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11968 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12027

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12028 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12087

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12088 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12147

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12148 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12207

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12208 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12267

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12268 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12327

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12328 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12387

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12388 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12447

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12448 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12507

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      12508 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12567
```

```
Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12568 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12627

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12628 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12687

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12688 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12747

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12748 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12807

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db      12808 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12867

Qy       2777 ------------------------------------------------------------ 2776

Db      12868 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12927

Qy       2777 ------------------------------------------------------------ 2776

Db      12928 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 12987

Qy       2777 ------------------------------------------------------------ 2776

Db      12988 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13047

Qy       2777 ------------------------------------------------------------ 2776

Db      13048 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13107

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||
Db      13108 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13167

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13168 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13227

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13228 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13287

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13288 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13347

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13348 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13407

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
             |||||||||||||||||||||||| |||
Db      13408 CTGGACGCGGATGAACAGGCAGACATCT 13435


RESULT 13
AY656167
LOCUS       AY656167               15287 bp    DNA     circular SYN 14-JAN-2005
DEFINITION  Chimeric dengue virus vector p4-D3L-ME, complete sequence.
ACCESSION   AY656167
VERSION     AY656167.1  GI:50345909
KEYWORDS    .
SOURCE      Chimeric dengue virus vector p4-D3L-ME
  ORGANISM  Chimeric dengue virus vector p4-D3L-ME
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 15287)
  AUTHORS   Blaney,J.E. Jr., Hanson,C.T., Firestone,C.Y., Hanley,K.A.,
            Murphy,B.R. and Whitehead,S.S.
  TITLE     Genetically modified, live attenuated dengue virus type 3 vaccine
            candidates
  JOURNAL   Am. J. Trop. Med. Hyg. 71 (6), 811-821 (2004)
   PUBMED   15642976
REFERENCE   2  (bases 1 to 15287)
  AUTHORS   Whitehead,S.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-2004) Laboratory of Infectious Diseases, NIAID,
            50 South Drive, Bldg 50, Room 6515, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .15287
                     /organism="Chimeric dengue virus vector p4-D3L-ME"
                     /mol_type="other DNA"
                     /db_xref="taxon:284997"
                     /focus
     source          1. .401
                     /organism="Dengue virus type 4"
                     /mol_type="other DNA"
                     /db_xref="taxon:11070"
     source          402. .2339
                     /organism="Dengue virus type 3"
                     /mol_type="other DNA"
                     /db_xref="taxon:11069"
                     /note="Sleman/78"
     source          2363. .10666
                     /organism="Dengue virus type 4"
                     /mol_type="other DNA"
                     /db_xref="taxon:11070"
     source          10676. .15037
                     /organism="Cloning vector pBR322"
                     /mol_type="other DNA"
                     /db_xref="taxon:47470"
     source          15038. .15287
                     /organism="Cloning vector pGEM-3"
                     /mol_type="other DNA"
                     /db_xref="taxon:90108"
     misc_feature    102. .10279
                     /note="encodes polyprotein precursor once XhoI linker
                     sequence is removed prior to preparation of infectious
                     transcripts corresponding to the viral genome"
     misc_feature    2340. .2362
                     /note="XhoI linker; required for stable replication in
                     Escherichia coli"
     misc_feature    10667. .10675
                     /note="KpnI/AgeI linker"
     promoter        15269. .15287
                     /note="SP6"
ORIGIN

  Query Match           34.0%;  Score 1636.8;  DB 11;  Length 15287;
  Best Local Similarity  86.4%;  Pred. No. 0;
  Matches 1943;  Conservative    0;  Mismatches   22;  Indels  283;  Gaps    1;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    | |      | ||||||||||||||||||||||||||||||
Db      11205 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11264

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11265 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11324

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11325 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11384

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11385 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11444

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11445 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11504
```

```
Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11505 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11564

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11565 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11624

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      11625 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11684

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11685 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11744

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11745 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11804

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11805 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11864

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11865 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11924

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGGACTGTTGGGCGCCATCTC 1888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11925 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGGACTGTTGGGCGCCATCTC 11984

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11985 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12044

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12045 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12104

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12105 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12164

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12165 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12224

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12225 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12284

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12285 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12344

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12345 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12404

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12405 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12464

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12465 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12524

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      12525 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12584

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12585 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12644

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12645 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12704

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12705 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12764

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12765 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12824

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12825 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12884

Qy       2777 ---------------------------------------------------------- 2776

Db      12885 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12944

Qy       2777 ---------------------------------------------------------- 2776

Db      12945 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13004

Qy       2777 ---------------------------------------------------------- 2776

Db      13005 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13064

Qy       2777 ---------------------------------------------------------- 2776

Db      13065 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13124

Qy       2777 -------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                               ||||||||||||||||||||||||||||
Db      13125 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13184

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13185 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13244

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13245 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13304

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13305 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13364

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13365 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13424

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
             ||||||||||||||||||||||  |||
Db      13425 CTGGACGCGGATGAACAGGCAGACATCT 13452
```

```
RESULT 14
AY028776/c
LOCUS       AY028776               13561 bp    DNA     circular SYN 22-DEC-2004
DEFINITION  TnphoZ mutagenesis vector pMHL120, complete sequence.
ACCESSION   AY028776
VERSION     AY028776.1  GI:13549398
KEYWORDS    .
SOURCE      TnphoZ mutagenesis vector pMHL120
  ORGANISM  TnphoZ mutagenesis vector pMHL120
            other sequences; artificial sequences; vectors.
REFERENCE   1  (bases 1 to 13561)
  AUTHORS   Lee,M.H., Nittayajarn,A., Ross,R.P., Rothschild,C.B., Parsonage,D.,
```

```
            Claiborne,A. and Rubens,C.E.
  TITLE     Characterization of Enterococcus faecalis alkaline phosphatase and
            use in identifying Streptococcus agalactiae secreted proteins
  JOURNAL   J. Bacteriol. 181 (18), 5790-5799 (1999)
   PUBMED   10482522
REFERENCE   2  (bases 1 to 13561)
  AUTHORS   Clancy,A., Lee,M.H., Jones,A.L. and Rubens,C.E.
  TITLE     Construction and characterization of transposon TnphoZ for the
            identification of genes encoding exported proteins in Streptococcus
            agalactiae
  JOURNAL   FEMS Microbiol. Lett. 241 (2), 257-264 (2004)
   PUBMED   15598541
REFERENCE   3  (bases 1 to 13561)
  AUTHORS   Lee,M.H., Clancy,A., Jones,A.L., Nittayajarn,A. and Rubens,C.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2001) Infectious Disease, Children's Hospital and
            Regional Medical Center, 4800 Sand Point Way NE, Seattle, WA 98105,
            USA
FEATURES             Location/Qualifiers
     source          1. .13561
                     /organism="TnphoZ mutagenesis vector pMHL120"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:156762"
                     /lab_host="Streptococcus agalactiae"
                     /note="transposon TnphoZ delivery vehicle"
     gene            complement(4167. .4928)
                     /gene="erm"
     CDS             complement(4167. .4928)
                     /gene="erm"
                     /note="confers erythromycin-resistance"
                     /codon_start=1
                     /transl_table=11
                     /product="rRNA methylase"
                     /protein_id="AAK27828.1"
                     /db_xref="GI:13549399"
                     /translation="MYHFRSDYMNKNIKYSQNFLTNEKVLNQIIKQLNLKETDTVYEI
                     GTGKGHLTTKLAKISKQVTSIELDSHLFNLSSEKLKLNIRVTLIHQDILQFQFPNKQR
                     YKIVGNIPYHLSTQIIKKVVFESHASDIYLIVEEGFYKRTLDIHRTLGLLLHTQVSIQ
                     QLLKLPAECFHPKPKVNSVLIKLTRHTTDVPDKYWKLYTYFVSKWVNREYRQLFTKNQ
                     FHQAMKHAKVNNLSTVTYEQVLSIFNSYLLFNGRK"
     gene            complement(6372. .6563)
                     /gene="rop"
     CDS             complement(6372. .6563)
                     /gene="rop"
                     /note="regulatory protein; involved in determination of
                     plasmid copy number"
                     /codon_start=1
                     /transl_table=11
                     /product="ROP protein"
                     /protein_id="AAK27829.1"
                     /db_xref="GI:13549400"
                     /translation="MTKQEKTALNMARFIRSQTLTLLEKLNELDADEQADICESLHDH
                     ADELYRSCLARFGDDGENL"
     gene            complement(7201. .8391)
                     /gene="tet"
     CDS             complement(7201. .8391)
                     /gene="tet"
                     /codon_start=1
                     /transl_table=11
                     /product="tetracycline-resistance protein"
                     /protein_id="AAK27830.1"
                     /db_xref="GI:13549401"
                     /translation="MKSNNALIVILGTVTLDAVGIGLVMPVLPGLLRDIVHSDSIASH
                     YGVLLALYALMQFLCAPVLGALSDRFGRRPVLLASLLGATIDYAIMATTPVLWILYAG
                     RIVAGITGATGAVAGAYIADITDGEDRARHFGLMSACFGVGMVAGPVAGGLLGAISLH
                     APFLAAAVLNGLNLLLGCFLMQESHKGERRPMPLRAFNPVSSFRWARGMTIVAALMTV
                     FFIMQLVGQVPAALWVIFGEDRFRWSATMIGLSLAVFGILHALAQAFVTGPATKRFGE
                     KQAIIAGMAADALGYVLLAFATRGWMAFPIMILLASGGIGMPALQAMLSRQVDDDHQG
                     QLQGSLAALTSLTSITGPLIVTAIYAASASTWNGLAWIVGAALYLVCLPALRRGAWSR
                     ATST"
     gene            complement(8683. .10032)
                     /gene="phoZ"
                     /allele="phoZ2"
                     /note="alkaline phosphatase; derived from Enterococcus
                     faecalis; expression is dependent upon integration event
                     to create in-frame start codon"
     gene            10619. .11413
                     /gene="aphA3"
     CDS             10619. .11413
                     /gene="aphA3"
                     /note="confers kanamycin-resistance"
                     /codon_start=1
                     /transl_table=11
                     /product="aminoglycoside phosphotransferase A3"
                     /protein_id="AAK27831.1"
                     /db_xref="GI:13549402"
                     /translation="MAKMRISPELKKLIEKYRCVKDTEGMSPAKVYKLVGENENLYLK
                     MTDSRYKGTTYDVEREKDMMLWLEGKLPVPKVLHFERHDGWSNLLMSEADGVLCSEEY
                     EDEQSPEKIIELYAECIRLFHSIDISDCPYTNSLDSRLAELDYLLNNDLADVDCENWE
                     EDTPFKDPRELYDFLKTEKPEEELVFSHGDLGDSNIFVKDGKVSGFIDLGRSGRADKW
                     YDIAFCVRSIREDIGEEQYVELFFDLLGIKPDWEKIKYYILLDELF"
     gene            complement(11817. .12515)
                     /gene="repAts"
     CDS             complement(11817. .12515)
                     /gene="repAts"
                     /codon_start=1
                     /transl_table=11
                     /product="temperature-sensitive RepA"
                     /protein_id="AAK27832.1"
                     /db_xref="GI:13549403"
                     /translation="MAIKNTKARNFGFLLYPDSIPNDWKEKLESLGVSMAVSPLHDMD
                     EKKDKDTWNNSNIIQNGKHYKKPHYHVIYIARNPVTIESVRNKIKRKLGNSSVAHVEI
                     LDYIKGSYEYLTHESKDAIAKNKHIYDKKDILNINDFDIDRYITLDESQKRELKNLLL
                     DIVDDYNLVNTKDLMAFIRLRGAEFGILNTNDVKDIVSTNSSAFRLWFEGNYQCGYRA
                     SYAKVLDAETGEIK"
ORIGIN

  Query Match             34.0%;  Score 1635.8;  DB 11;  Length 13561;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 1637;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1140 ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT 1199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8667 ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT 8608

Qy       1200 AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCT 1259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8607 AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCT 8548

Qy       1260 AAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC 1319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8547 AAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTC 8488

Qy       1320 GTCTTCAAGAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTA 1379
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8487 GTCTTCAAGAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTA 8428

Qy       1380 TCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATC 1439
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8427 TCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATC 8368

Qy       1440 GTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTG 1499
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8367 GTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTG 8308

Qy       1500 CCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTG 1559
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8307 CCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTG 8248

Qy       1560 CTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGAC 1619
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8247 CTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGAC 8188

Qy       1620 CGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCG 1679
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8187 CGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCG 8128
```

```
Qy       1680 ATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATC 1739
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8127 ATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATC 8068

Qy       1740 ACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGAT 1799
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8067 ACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGAT 8008

Qy       1800 CGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCC 1859
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8007 CGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCC 7948

Qy       1860 GTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTG 1919
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7947 GTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTG 7888

Qy       1920 CTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAG 1979
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7887 CTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAG 7828

Qy       1980 CGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGC 2039
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7827 CGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGC 7768

Qy       2040 ATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTG 2099
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7767 ATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTG 7708

Qy       2100 CCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATC 2159
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7707 CCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATC 7648

Qy       2160 GGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGT 2219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7647 GGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGT 7588

Qy       2220 CCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCG 2279
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7587 CCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCG 7528

Qy       2280 CTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATT 2339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7527 CTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATT 7468

Qy       2340 CTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTA 2399
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7467 CTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTA 7408

Qy       2400 GATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCG 2459
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7407 GATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCG 7348

Qy       2460 ATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGG 2519
             |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7347 ATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGG 7288

Qy       2520 TTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGT 2579
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7287 TTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGT 7228

Qy       2580 GCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCA 2639
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7227 GCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCA 7168

Qy       2640 CCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACC 2699
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7167 CCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACC 7108

Qy       2700 CTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGG 2759
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7107 CTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGG 7048

Qy       2760 GCAGCGTTGGGTCCTGGGC 2778
              ||||||||||||||||| |
Db       7047 GCAGCGTTGGGTCCTGGCC 7029
```

RESULT 15
SYNCLCTAPS
LOCUS SYNCLCTAPS 6343 bp DNA circular SYN 27-APR-1993
DEFINITION Expression cloning vector pNP6CRMCTAP colicin E1-connective tissue activating peptide fusion protein gene, complete cds.
ACCESSION M88538
VERSION M88538.1 GI:208128
KEYWORDS colicin E1; colicin:E1-CTAP-III(Leu21) fusion protein; connective tissue activating peptide-III; expression vector.
SOURCE unidentified cloning vector
ORGANISM unidentified cloning vector
other sequences; artificial sequences; vectors.
REFERENCE 1 (bases 1 to 6343)
AUTHORS Waleh,N.S., Sohel,I., Lazar,J.B., Hudson,D.V., Sze,P., Underhill,P.A. and Johnson,P.H.
TITLE Expression of synthetic genes encoding fused proteins under tight control of modified regulatory regions of the colicin operon
JOURNAL Gene 117 (1), 7-14 (1992)
PUBMED 1644316
COMMENT Original source text: Cloning vector DNA.
FEATURES Location/Qualifiers

source	1. .6343
	/organism="unidentified cloning vector"
	/mol_type="genomic DNA"
	/db_xref="taxon:45196"
-35_signal	5794. .5799
-10_signal	5817. .5822
promoter	5829. .5861
RBS	5895. .5899
	/evidence=experimental
CDS	5907. .6182
	/codon_start=1
	/transl_table=11
	/product="Col-CTAP-III(Leu21) small fusion protein"
	/protein_id="AAA73219.1"
	/db_xref="GI:208129"
	/translation="METAVMNLAKGKEESLDSDLYAELRCLCIKTTSGIHPKNIQSLE VIGKGTHCNQVEVIATLKDGRKICLDPDAPRIKKIVQKKLAGDESAD"
sig_peptide	5907. .5924
	/note="leader sequence; putative"
mat_peptide	5925. .6179
	/product="connective tissue activating peptide III"
	/standard_name="CTAP-III"
	/note="(Leu21); CTAP-III(Leu21); putative"
terminator	6278. .6308
	/standard_name="Rho-dependent terminator"

ORIGIN

Query Match 33.7%; Score 1624; DB 11; Length 6343;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1650; Conservative 0; Mismatches 20; Indels 4; Gaps 1;

```
Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
              ||| ||| ||  | |     |  |      |  ||||||||||||||||||||||||||
Db        529 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 588

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 648

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 708

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAG----AATTCTCATGTTTGAC 1344
              ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        709 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTAATTCTCATGTTTGAC 768
```

```
Qy       1345 AGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTC 1404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        769 AGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTC 828

Qy       1405 AGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGG 1464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 AGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGG 888

Qy       1465 ATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCC 1524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 ATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCC 948

Qy       1525 ATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAAT 1584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 ATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAAT 1008

Qy       1585 TTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGC 1644
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1009 TTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGC 1068

Qy       1645 TCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGT 1704
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 TCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGT 1128

Qy       1705 GGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTG 1764
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 GGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTG 1188

Qy       1765 GCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGA 1824
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1189 GCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGA 1248

Qy       1825 GCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCA 1884
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1249 GCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCA 1308

Qy       1885 TCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGG 1944
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1309 TCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGG 1368

Qy       1945 GCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCT 2004
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1369 GCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCT 1428

Qy       2005 TCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGA 2064
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1429 TCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGA 1488

Qy       2065 CTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCG 2124
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1489 CTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCG 1548

Qy       2125 GCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAA 2184
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1549 GCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAA 1608

Qy       2185 TCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGA 2244
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1609 TCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGA 1668

Qy       2245 AGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCG 2304
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1669 AGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCG 1728

Qy       2305 CGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGA 2364
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1729 CGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGA 1788

Qy       2365 TGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTC 2424
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1789 TGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTC 1848

Qy       2425 AAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGG 2484
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1849 AAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGG 1908

Qy       2485 CGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCC 2544
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1909 CGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCC 1968

Qy       2545 TATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCT 2604
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1969 TATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCT 2028

Qy       2605 GAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCA 2664
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2029 GAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCA 2088

Qy       2665 ATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTC 2724
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2089 ATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTC 2148

Qy       2725 CGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGGC 2778
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       2149 CGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCC 2202
```

Search completed: January 18, 2006, 04:26:11
Job time : 24858 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 17:02:23 ; Search time 2993.54 Seconds
(without alignments)
10717.687 Million cell updates/sec

Title: US-10-511-327-7
Perfect score: 4814
Sequence: 1 ggtaccctgtgaatgcgcaa.....taagatgctttctgtgact 4814

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4814	100.0	4814	10	ACF58174	Acf58174 Modified
	2	1636.8	34.0	15153	12	ADG93315	Adg93315 DEN3 (Sle
	3	1636.8	34.0	15159	12	ADG93313	Adg93313 DEN2 (Ton
	4	1560.6	32.4	7311	14	AEA06371	Aea06371 Novel bac
c	5	1477.2	30.7	5926	2	AAV32978	Aav32978 Tn7 donor
c	6	1477.2	30.7	5926	6	AAD45060	Aad45060 pEM delta
c	7	1476	30.7	5926	2	AAV32977	Aav32977 Tn7 donor
c	8	1476	30.7	5926	6	AAD45059	Aad45059 Transposo
c	9	1476	30.7	5926	12	ADG46817	Adg46817 Donor pla
c	10	1474	30.6	5926	12	ADG46818	Adg46818 pEM DNA (
	11	1447.8	30.1	4840	4	AAF59457	Aaf59457 Npro-hGH
c	12	1446.8	30.1	3474	2	AAQ29146	Aaq29146 pmTNF-MPH
c	13	1446.8	30.1	3474	2	AAQ51546	Aaq51546 Expressio
c	14	1446.8	30.1	3801	12	ADL72227	Adl72227 DNA seque
	15	1446.8	30.1	4245	8	ABZ70623	Abz70623 Plasmid p
c	16	1446.8	30.1	5201	12	ADL72228	Adl72228 DNA seque
	17	1446.8	30.1	11184	12	ADP86274	Adp86274 Hepatitis
	18	1446.8	30.1	11184	12	ADP86276	Adp86276 Hepatitis
	19	1446.8	30.1	11184	12	ADP86277	Adp86277 Hepatitis
	20	1446.8	30.1	11313	12	ADP86273	Adp86273 Hepatitis
	21	1446.8	30.1	11313	12	ADP86264	Adp86264 Hepatitis
	22	1446.8	30.1	11313	12	ADP86266	Adp86266 Hepatitis
	23	1446.8	30.1	11313	12	ADP86265	Adp86265 Hepatitis
	24	1446.8	30.1	11313	12	ADP86268	Adp86268 Hepatitis
	25	1446.8	30.1	11313	12	ADP86270	Adp86270 Hepatitis
	26	1446.8	30.1	11313	12	ADP86271	Adp86271 Hepatitis
	27	1446.8	30.1	11313	12	ADP86272	Adp86272 Hepatitis
	28	1446.8	30.1	11313	12	ADP86269	Adp86269 Hepatitis
	29	1446.8	30.1	11313	12	ADP86275	Adp86275 Hepatitis
	30	1446.8	30.1	11313	12	ADP86267	Adp86267 Hepatitis
	31	1446.8	30.1	12305	13	ADT77488	Adt77488 HCV BB7 c
	32	1446.8	30.1	12306	10	ADI41414	Adi41414 BB7 nucle
	33	1446.8	30.1	12311	13	ADT77500	Adt77500 reporter
	34	1446.8	30.1	12315	10	ADI41413	Adi41413 BB7M4hRLu
	35	1446.8	30.1	12980	2	AAV59364	Aav59364 Hepatitis
	36	1446.8	30.1	12980	6	ABK87286	Abk87286 Hepatitis
	37	1446.8	30.1	12980	8	ACA62469	Aca62469 DNA encod
c	38	1445.2	30.0	3423	2	AAQ11083	Aaq11083 Plasmid p
c	39	1445.2	30.0	5201	12	ADL72229	Adl72229 DNA seque
	40	1444.8	30.0	2315	10	ADF44791	Adf44791 De novo s
	41	1444.8	30.0	2600	10	ADF44790	Adf44790 De novo s
c	42	1444.8	30.0	4009	2	AAQ70219	Aaq70219 Plasmid p
c	43	1442.8	30.0	3880	4	AAF25570	Aaf25570 Vector pS
	44	1442.8	30.0	5115	2	AAV81319	Aav81319 Plasmid p
	45	1442.8	30.0	5115	5	AAH74745	Aah74745 Nucleotid

ALIGNMENTS

RESULT 1
ACF58174
ID ACF58174 standard; DNA; 4814 BP.
XX
AC ACF58174;
XX
DT 15-JAN-2004 (first entry)
XX
DE Modified Tn7-Kan(Stop)-transposon.
XX
KW Transposon; genetic engineering; transposase; Cat-Mu; Tn7-Kan; ds.
XX
OS Synthetic.
XX
PN WO2003087370-A1.
XX
PD 23-OCT-2003.
XX
PF 14-APR-2003; 2003WO-FI000285.
XX
PR 18-APR-2002; 2002FI-00000746.
XX
PA (FINN-) FINNZYMES OY.
XX
PI Savilahti H, Tieaho V;
XX
DR WPI; 2003-845329/78.
XX
PT New transposon nucleic acid comprising a genetically engineered
PT translation stop signal within a transposon end sequence recognized by a
PT transposase useful for producing deletion derivatives of polypeptide.
XX
PS Claim 7; Page 31-32; 0pp; English.
XX
CC The invention relates to a transposon nucleic acid comprising a
CC genetically engineered translation stop signal in the three reading
CC frames at least partly within a transposon end sequence recognized by a
CC transposase. The transposon is useful for producing deletion derivatives
CC of polypeptide coding nucleic acids. The method involves performing a
CC transposition reaction in the presence of a target nucleic acid
CC containing a polypeptide coding nucleic acid and in the presence of a
CC transposon containing a genetically engineered translation stop signal

CC sequence in three reading frames at least partly within a transposon end
CC sequence recognized by a transposase; and recovering a target nucleic
CC acid having the transposon incorporated in the protein coding nucleic
CC acid. The present sequence represents a modified Tn7-Kan(Stop)-
CC transposon
XX
SQ Sequence 4814 BP; 1214 A; 1206 C; 1137 G; 1257 T; 0 U; 0 Other;

Query Match 100.0%; Score 4814; DB 10; Length 4814;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4814; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGTACCCTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGTACCCTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCT 60

Qy         61 CCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGA 120

Qy        121 TCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGA 180

Qy        181 ATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCT 240

Qy        241 GAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGT 300

Qy        301 CAGCGCCCTGCACCATTATGTTCCGGATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CAGCGCCCTGCACCATTATGTTCCGGATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAA 360

Qy        361 TGGCAGATCCCTGGCTTGTTGTCCACAACCGTTAAACCTTAAAAGCTTTAAAAGCCTTAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGGCAGATCCCTGGCTTGTTGTCCACAACCGTTAAACCTTAAAAGCTTTAAAAGCCTTAT 420

Qy        421 ATATTCTTTTTTTTCTTATAAAACTTAAAACCTTAGAGGCTATTTAAGTTGCTGATTTAT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATATTCTTTTTTTTCTTATAAAACTTAAAACCTTAGAGGCTATTTAAGTTGCTGATTTAT 480

Qy        481 ATTAATTTTATTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATTAATTTTATTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTT 540

Qy        541 AGCCATGAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTAAACATGAGAGCTTAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGCCATGAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTAAACATGAGAGCTTAG 600

Qy        601 TACGTTAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TACGTTAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAG 660

Qy        661 GTTGAACTGCTGATCTTCGGATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGCAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTTGAACTGCTGATCTTCGGATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGCAG 720

Qy        721 ATCCCTGGCTTGTTGTCCACAACCGTTAAACCTTAAAAGCTTTAAAAGCCTTATATATTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 ATCCCTGGCTTGTTGTCCACAACCGTTAAACCTTAAAAGCTTTAAAAGCCTTATATATTC 780

Qy        781 TTTTTTTTCTTATAAAACTTAAAACCTTAGAGGCTATTTAAGTTGCTGATTTATATTAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTTTTTTCTTATAAAACTTAAAACCTTAGAGGCTATTTAAGTTGCTGATTTATATTAAT 840

Qy        841 TTTATTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTTAGCCAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TTTATTGTTCAAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTTAGCCAT 900

Qy        901 GAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTAAACATGAGAGCTTAGTACGTT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAGAGCTTAGTACGTTAGCCATGAGGGTTTAGTTCGTTAAACATGAGAGCTTAGTACGTT 960

Qy        961 AAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAGGTTGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAACATGAGAGCTTAGTACGTGAAACATGAGAGCTTAGTACGTACTATCAACAGGTTGAA 1020

Qy       1021 CTGCTGATCTTCGGATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGCATCCGGAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTGCTGATCTTCGGATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGCATCCGGAT 1080

Qy       1081 CTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCA 1140

Qy       1141 TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA 1200

Qy       1201 GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTA 1260

Qy       1261 AGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG 1320

Qy       1321 TCTTCAAGAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTAT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TCTTCAAGAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTAT 1380

Qy       1381 CACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCG 1440

Qy       1441 TCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGC 1500

Qy       1501 CGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGC 1560

Qy       1561 TAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACC 1620

Qy       1621 GCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 GCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGA 1680

Qy       1681 TCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCA 1740

Qy       1741 CCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATC 1800

Qy       1801 GGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCG 1860

Qy       1861 TGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGC 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGC 1920

Qy       1921 TCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGC 1980

Qy       1981 GTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCA 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1981 GTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCA 2040
Qy       2041 TGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGC 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGC 2100
Qy       2101 CGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 CGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCG 2160
Qy       2161 GCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTC 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTC 2220
Qy       2221 CCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGC 2280
Qy       2281 TGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTC 2340
Qy       2341 TTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAG 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAG 2400
Qy       2401 ATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGA 2460
Qy       2461 TCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGT 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGT 2520
Qy       2521 TGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTG 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 TGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTG 2580
Qy       2581 CATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCAC 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 CATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCAC 2640
Qy       2641 CACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCC 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCC 2700
Qy       2701 TTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 TTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGG 2760
Qy       2761 CAGCGTTGGGTCCTGGGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATC 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 CAGCGTTGGGTCCTGGGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATC 2820
Qy       2821 CATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTA 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 CATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCATGTTCATCATCAGTA 2880
Qy       2881 ACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 ACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACCCCCATGAACAGAAAT 2940
Qy       2941 CCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGC 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCCCTTAACATGGCCCGC 3000
Qy       3001 TTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAA 3060
Qy       3061 CAGGCAGAGCTCTTACTGTCATGCCATCCGTATGAGTGAGTAGAATAAAGTCTTAAACTG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 CAGGCAGAGCTCTTACTGTCATGCCATCCGTATGAGTGAGTAGAATAAAGTCTTAAACTG 3120

Qy       3121 AACAAAATAGATCTAAACTATGACAATAAAGTCTTAAACTAGACAGAATAGTTGTAAACT 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 AACAAAATAGATCTAAACTATGACAATAAAGTCTTAAACTAGACAGAATAGTTGTAAACT 3180
Qy       3181 GAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAA 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 GAAATCAGTCCAGTTATGCTGTGAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAA 3240
Qy       3241 CTCTTCATTTTCTGAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGGGTCGACGC 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 CTCTTCATTTTCTGAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGGGTCGACGC 3300
Qy       3301 GGCCGCTAACTATAACGGTCCTAAGGTAGCGAGTTTAAACGATATCGGATCCGGCCGCCG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 GGCCGCTAACTATAACGGTCCTAAGGTAGCGAGTTTAAACGATATCGGATCCGGCCGCCG 3360
Qy       3361 CTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCATC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 CTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTGACTCATACCAGGCCTGAATCGCCCCATC 3420
Qy       3421 ATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTT 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 ATCCAGCCAGAAAGTGAGGGAGCCACGGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTT 3480
Qy       3481 GGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGAAGATGCGTGAT 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 GGTGATTTTGAACTTTTGCTTTGCCACGGAACGGTCTGCGTTGTCGGGAAGATGCGTGAT 3540
Qy       3541 CTGATCCTTCAACTCAGCAAGAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCAAGTC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTGATCCTTCAACTCAGCAAGAGTTCGATTTATTCAACAAAGCCGCCGTCCCGTCAAGTC 3600
Qy       3601 AGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCG 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 AGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTCTGATTAGAAAAACTCATCG 3660
Qy       3661 AGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAA 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 AGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAA 3720
Qy       3721 AGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCC 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 AGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCC 3780
Qy       3781 TGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCG 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 TGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCG 3840
Qy       3841 TCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAAT 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 TCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAAT 3900
Qy       3901 GGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCA 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 GGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCA 3960
Qy       3961 TCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGA 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 TCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGA 4020
Qy       4021 AATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGG 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 AATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGG 4080
Qy       4081 AACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 AACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGG 4140
Qy       4141 AATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATA 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATA 4200
```

```
Qy       4201 AAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCA 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 AAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCA 4260

Qy       4261 TCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCG 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 TCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCG 4320

Qy       4321 GGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCAT 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCAT 4380

Qy       4381 TTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTT 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 TTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTT 4440

Qy       4441 TCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTT 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 TCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTT 4500

Qy       4501 ATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 ATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACA 4560

Qy       4561 ACGTGGCTTACTAGGATCCGATATCATTTAAATCTAGGGATAACAGGGTAATACTAGTGT 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 ACGTGGCTTACTAGGATCCGATATCATTTAAATCTAGGGATAACAGGGTAATACTAGTGT 4620

Qy       4621 CGACCAACCAGATAAGTGAAATCTAGTTCCAAACTATTTTGTCATTTTTAATTTTCGTAT 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 CGACCAACCAGATAAGTGAAATCTAGTTCCAAACTATTTTGTCATTTTTAATTTTCGTAT 4680

Qy       4681 TAGCTTACGACGCTACACCCAGTTCCCATCTATTTTGTCACTCTTCCCTAAATAATCCTT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 TAGCTTACGACGCTACACCCAGTTCCCATCTATTTTGTCACTCTTCCCTAAATAATCCTT 4740

Qy       4741 AAAAACTCCATTTCCACCCCTCCCAGTTCCCAACTATTTTCTACTCACTCACCGTAAGAT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 AAAAACTCCATTTCCACCCCTCCCAGTTCCCAACTATTTTCTACTCACTCACCGTAAGAT 4800

Qy       4801 GCTTTTCTGTGACT 4814
             ||||||||||||||
Db       4801 GCTTTTCTGTGACT 4814
```

RESULT 2
ADG93315
ID ADG93315 standard; DNA; 15153 BP.
XX
AC ADG93315;
XX
DT 11-MAR-2004 (first entry)
XX
DE DEN3 (Sleman/78) cDNA plasmid P3.
XX
KW immunogenic composition; dengue type 1; dengue type 2; dengue type 3;
KW dengue type 4; virucide; immunostimulant; vaccine; tetravalent vaccine;
KW dengue virus; delta30; attenuating mutation; humoral response;
KW cellular response; non-structural protein; structural protein;
KW dengue virus serotype; gene; ds; plasmid P3.
XX
OS Dengue virus type 3.
XX
PN WO2003092592-A2.
XX
PD 13-NOV-2003.
XX
PF 25-APR-2003; 2003WO-US013279.
XX
PR 03-MAY-2002; 2002US-0377860P.
PR 23-DEC-2002; 2002US-0436500P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Whitehead SS, Murphy BR, Markoff L, Falgout B, Blaney J;
PI Hanley K;
XX
DR WPI; 2004-022612/02.
DR P-PSDB; ADG93316.
XX
PT New tetravalent vaccine containing a common nucleotide deletion in the 3'
PT untranslated region of dengue types 1, 2, 3, and 4, useful for preventing
PT of disease in humans caused by dengue virus, or for inducing immune
PT response.
XX
PS Disclosure; SEQ ID NO 48; 181pp; English.
XX
CC This invention relates to a novel immunogenic composition being
CC tetravalent and containing a common nucleotide deletion in the 3'
CC untranslated region of dengue types 1, 2, 3, and 4. The invention may be
CC useful for the development of compounds with a virucide or
CC immunostimulant activity or as a vaccine. The tetravalent vaccine is
CC useful in the prevention of disease in humans caused by dengue virus.
CC Unlike previous tetravalent vaccine, the new tetravalent vaccine is
CC unique since they contain a common shared attenuating mutation which
CC eliminates the possibility of generating a virulent wild type virus in a
CC subject to be vaccinated since each component of the vaccine possesses
CC the same delta30 attenuating deletion mutation. The vaccine also is able
CC to induce humoral and cellular responses against al of the (non-
CC)structural proteins present in each dengue virus serotype. The present
CC sequence is that of the DEN3 cDNA plasmid P3 which is related to the
CC invention.
XX
SQ Sequence 15153 BP; 4434 A; 3440 C; 3943 G; 3336 T; 0 U; 0 Other;

Query Match 34.0%; Score 1636.8; DB 12; Length 15153;
Best Local Similarity 86.4%; Pred. No. 3.1e-297;
Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1;

```
Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |   |  |    |  |||||||||||||||||||||||||||||||
Db      11246 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11305

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11306 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11365

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11366 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11425

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11426 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11485

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11486 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11545

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11546 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11605

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11606 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11665

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      11666 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11725

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11726 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11785

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
```

```
Db      11786 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11845

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768

Db      11846 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11905

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828

Db      11906 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11965

Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888

Db      11966 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12025

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948

Db      12026 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12085

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008

Db      12086 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12145

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068

Db      12146 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12205

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128

Db      12206 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12265

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188

Db      12266 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12325

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248

Db      12326 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12385

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308

Db      12386 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12445

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368

Db      12446 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12505

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428

Db      12506 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12565

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488

Db      12566 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12625

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548

Db      12626 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12685

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608

Db      12686 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12745

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668

Db      12746 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12805

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728

Db      12806 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12865

Qy       2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG---------- 2776

Db      12866 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12925

Qy       2777 ---------------------------------------------------------- 2776

Db      12926 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12985

Qy       2777 ---------------------------------------------------------- 2776

Db      12986 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13045

Qy       2777 ---------------------------------------------------------- 2776

Db      13046 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13105

Qy       2777 ---------------------------------------------------------- 2776

Db      13106 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13165

Qy       2777 ----------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805

Db      13166 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13225

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865

Db      13226 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13285

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925

Db      13286 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13345

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985

Db      13346 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13405

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045

Db      13406 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13465

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073

Db      13466 CTGGACGCGGATGAACAGGCAGACATCT 13493
```

RESULT 3
ADG93313
ID ADG93313 standard; DNA; 15159 BP.
XX
AC ADG93313;
XX
DT 11-MAR-2004 (first entry)
XX
DE DEN2 (Tonga/74) cDNA plasmid P2.
XX
KW immunogenic composition; dengue type 1; dengue type 2; dengue type 3;
KW dengue type 4; virucide; immunostimulant; vaccine; tetravalent vaccine;
KW dengue virus; delta30; attenuating mutation; humoral response;
KW cellular response; non-structural protein; structural protein;
KW dengue virus serotype; gene; ds; plasmid P2.
XX
OS Dengue virus type 2.
XX
PN WO2003092592-A2.
XX
PD 13-NOV-2003.
XX
PF 25-APR-2003; 2003WO-US013279.
XX
PR 03-MAY-2002; 2002US-0377860P.
PR 23-DEC-2002; 2002US-0436500P.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Whitehead SS, Murphy BR, Markoff L, Falgout B, Blaney J;

PI Hanley K;
XX
DR WPI; 2004-022612/02.
DR P-PSDB; ADG93314.
XX
PT New tetravalent vaccine containing a common nucleotide deletion in the 3' untranslated region of dengue types 1, 2, 3, and 4, useful for preventing of disease in humans caused by dengue virus, or for inducing immune response.
XX
PS Disclosure; SEQ ID NO 46; 181pp; English.
XX
CC This invention relates to a novel immunogenic composition being tetravalent and containing a common nucleotide deletion in the 3' untranslated region of dengue types 1, 2, 3, and 4. The invention may be useful for the development of compounds with a virucide or immunostimulant activity or as a vaccine. The tetravalent vaccine is useful in the prevention of disease in humans caused by dengue virus. Unlike previous tetravalent vaccine, the new tetravalent vaccine is unique since they contain a common shared attenuating mutation which eliminates the possibility of generating a virulent wild type virus in a subject to be vaccinated since each component of the vaccine possesses the same delta30 attenuating deletion mutation. The vaccine also is able to induce humoral and cellular responses against al of the (non-)structural proteins present in each dengue virus serotype. The present sequence is that of the DEN2 cDNA plasmid P2 which is related to the invention.
XX
SQ Sequence 15159 BP; 4518 A; 3436 C; 3896 G; 3309 T; 0 U; 0 Other;

Query Match 34.0%; Score 1636.8; DB 12; Length 15159;
Best Local Similarity 86.4%; Pred. No. 3.1e-297;
Matches 1943; Conservative 0; Mismatches 22; Indels 283; Gaps 1;

```
Qy      1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
             ||| ||| ||  | |    |  |     |  ||||||||||||||||||||||||||||
Db     11252 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 11311

Qy      1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11312 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 11371

Qy      1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11372 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 11431

Qy      1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11432 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 11491

Qy      1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11492 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 11551

Qy      1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11552 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 11611

Qy      1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11612 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 11671

Qy      1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     11672 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTGGCGCTATATGCGTTGATGCAATTTCT 11731

Qy      1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11732 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 11791

Qy      1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11792 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 11851

Qy      1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11852 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 11911

Qy      1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11912 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 11971

Qy      1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11972 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 12031

Qy      1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12032 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 12091

Qy      1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12092 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 12151

Qy      2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12152 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 12211

Qy      2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12212 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 12271

Qy      2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12272 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 12331

Qy      2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12332 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 12391

Qy      2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12392 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 12451

Qy      2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12452 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 12511

Qy      2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12512 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 12571

Qy      2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     12572 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 12631

Qy      2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12632 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 12691

Qy      2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12692 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 12751

Qy      2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12752 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 12811

Qy      2669 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 2728
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12812 TTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC 12871

Qy      2729 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGG------------ 2776
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12872 ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGC 12931
```

```
Qy       2777 ------------------------------------------------------------ 2776

Db      12932 ATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTAG 12991

Qy       2777 ------------------------------------------------------------ 2776

Db      12992 CAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCG 13051

Qy       2777 ------------------------------------------------------------ 2776

Db      13052 ACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGG 13111

Qy       2777 ------------------------------------------------------------ 2776

Db      13112 AAGTCAGCGCCCTGCACCATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCC 13171

Qy       2777 ------------------------------GCTGGCATTGACCCTGAGTGATTTTTCTC 2805
                                            ||||||||||||||||||||||||||||
Db      13172 TGTGGAACACCTACATCTGTATTAACGAAGCGCTGGCATTGACCCTGAGTGATTTTTCTC 13231

Qy       2806 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 2865
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13232 TGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCACAACGTTCCAGTAACCGGGCA 13291

Qy       2866 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 2925
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13292 TGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCGGTATCATTACC 13351

Qy       2926 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 2985
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13352 CCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGCC 13411

Qy       2986 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 3045
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      13412 CTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAG 13471

Qy       3046 CTGGACGCGGATGAACAGGCAGAGCTCT 3073
              |||||||||||||||||||||||| |||
Db      13472 CTGGACGCGGATGAACAGGCAGACATCT 13499
```

RESULT 4
AEA06371
ID AEA06371 standard; DNA; 7311 BP.
XX
AC AEA06371;
XX
DT 14-JUL-2005 (first entry)
XX
DE Novel bacterial host cell-related plasmid pCLO43 DNA SeqID7.
XX
KW gene expression; protein production; chromosome; genome; cell culture;
KW antibiotic-resistance; ds.
XX
OS Unidentified.
XX
PN WO2005042750-A1.
XX
PD 12-MAY-2005.
XX
PF 29-OCT-2004; 2004WO-DK000750.
XX
PR 31-OCT-2003; 2003DK-00001624.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Olsen C, Rasmussen MD;
XX
DR WPI; 2005-346878/35.
XX
PT New bacterial host cell comprising at least two copies of an
PT amplification unit in its genome, useful for increased production of a
PT protein with several copies of a gene of interest stably chromosomally
PT integrated.
XX
PS Example; SEQ ID NO 7; 53pp; English.
XX
CC This invention relates to a novel bacterial host cell comprising at least
CC two copies of an amplification unit in its genome, where two or more
CC amplified chromosomal copies of the gene of interest are produced. The
CC amplification unit comprises at least one copy of a gene of interest and
CC an expressible conditionally essential gene, where the conditionally
CC essential gene is either promoterless or transcribed from a heterologous
CC promoter having an activity substantially lower than the endogenous
CC promoter of the conditionally essential gene, and where the conditionally
CC essential gene if not functional would render the cell auxotrophic for at
CC least one specific substance or unable to utilize one or more specific
CC sole carbon source.The bacterial host cell is useful for producing of a
CC protein having several copies of a gene of interest stably chromosomally
CC integrated, without leaving antibiotic resistance marker genes in the
CC strains. The present sequence is that of a plasmid DNA sequence which was
CC used in the exemplification of the invention.
XX
SQ Sequence 7311 BP; 1926 A; 1706 C; 1736 G; 1943 T; 0 U; 0 Other;

```
  Query Match             32.4%;  Score 1560.6;  DB 14;  Length 7311;
  Best Local Similarity   98.8%;  Pred. No. 5.7e-283;
  Matches 1572;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
              ||| ||| ||  | |     |  |     |  ||||||||||||||||||||||||||||
Db       5721 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT 5780

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5781 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 5840

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5841 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 5900

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 1348
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5901 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATTCTCATGTTTGACAGCT 5960

Qy       1349 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 1408
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5961 TATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGC 6020

Qy       1409 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 1468
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6021 ACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGC 6080

Qy       1469 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 1528
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6081 TGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTC 6140

Qy       1529 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 1588
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6141 CGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCT 6200

Qy       1589 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 1648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6201 ATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGC 6260

Qy       1649 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 1708
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6261 TTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGAT 6320

Qy       1709 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 1768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6321 CCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGC 6380

Qy       1769 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6381 CTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGC 6440
```

```
Qy       1829 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 1888
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6441 TTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC 6500

Qy       1889 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 1948
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6501 CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTG 6560

Qy       1949 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 2008
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6561 CTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAA 6620

Qy       2009 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 2068
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6621 CCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGT 6680

Qy       2069 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 2128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6681 CTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGA 6740

Qy       2129 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 2188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6741 GGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT 6800

Qy       2189 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 2248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6801 GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCA 6860

Qy       2249 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 2308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6861 GGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGAC 6920

Qy       2309 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 2368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6921 GCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC 6980

Qy       2369 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 2428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6981 CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGG 7040

Qy       2429 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGAT 2488
             |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       7041 ATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGAT 7100

Qy       2489 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 2548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7101 TTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA 7160

Qy       2549 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 2608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7161 CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAAT 7220

Qy       2609 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 2668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7221 GGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTC 7280

Qy       2669 TTGCGGAGAACTGTGAATGCGCAAACCAACC 2699
             ||||||||||||||||||||||||||||||
Db       7281 TTGCGGAGAACTGTGAATGCGCAAACCAACC 7311
```

```
RESULT 5
AAV32978/c
ID   AAV32978 standard; DNA; 5926 BP.
XX
AC   AAV32978;
XX
DT   17-NOV-1998  (first entry)
XX
DE   Tn7 donor plasmid sequence.
XX
KW   Tn7; donor plasmid; transposon; ATP-utilising regulatory protein;
KW   transposable element; DNA sequencing; genetic analysis;
KW   insertional mutagenesis; ss.
XX
OS   Synthetic.
XX
PN   WO9837205-A1.
XX
PD   27-AUG-1998.
XX
PF   20-FEB-1998;   98WO-US003353.
XX
PR   20-FEB-1997;   97US-00037955.
XX
PA   (CRAI/) CRAIG N L.
XX
PI   Craig NL;
XX
DR   WPI; 1998-467567/40.
XX
PT   Transposon(s) encoding mutant ATP using proteins for insertion - which is
PT   efficient and random, with reduced site specificity; for DNA sequencing
PT   and altering gene expression.
XX
PS   Disclosure; Fig 10C; 143pp; English.
XX
CC   The present sequence represents a Tn7 donor plasmid pEM delta. Transposon
CC   Tn7 encodes an ATP-utilising regulatory protein that contains a mutation
CC   that allows efficient and simple insertion of, and reduced target site
CC   specificity of, a transposable element derived from the transposon. The
CC   ATP-regulatory protein and compositions are used for the efficient, non-
CC   specific and simple insertion of a transposon or transposable element
CC   into a DNA segment. This is useful in DNA sequencing, for genetic
CC   analysis by insertional mutagenesis, or for alteration of gene expression
CC   by insertion of a desired sequence
XX
SQ   Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

  Query Match             30.7%;  Score 1477.2;  DB 2;  Length 5926;
  Best Local Similarity   93.4%;  Pred. No. 2.4e-267;
  Matches 1594;  Conservative    0;  Mismatches   38;  Indels   74;  Gaps    2;
```

```
Qy       3087 TCCGTATGAGTGAGTAGAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA 3146
             ||   ||| | | | | ||||||||||||||||||||||||||||||||||||||||||||||
Db       5568 TCTAAATGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA 5509

Qy       3147 TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAA 3206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5508 TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAA 5449

Qy       3207 AAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTG 3266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5448 AAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTG 5389

Qy       3267 CCCGTCGTATTAAAGAGGGGCGTGGGGTCGACGCGGCCGCTAACTATAACGGTCCTAAGG 3326
             ||||||||||||||||||||||||||||||||
Db       5388 CCCGTCGTATTAAAGAGGGGCGTGGGGTCGAC---------------------------- 5357

Qy       3327 TAGCGAGTTTAAACGATATCGGATCCGGCCGCCGCTGAGGTCTGCCTCGTGAAGAAGGTG 3386
             |    ||    ||  | ||||||||||||||||||||||||||||||||||||||||||||||
Db       5356 --------------CTGCAGGGGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTG 5312

Qy       3387 TTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAC 3446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5311 TTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAC 5252

Qy       3447 GGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCA 3506
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5251 GGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCA 5192

Qy       3507 CGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAGAGTTC 3566
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       5191 CGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTC 5132
```

```
Qy       3567 GATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAA 3626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5131 GATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAA 5072

Qy       3627 CCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT 3686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5071 CCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT 5012

Qy       3687 CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAA 3746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5011 CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAA 4952

Qy       3747 CTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 3806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4951 CTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 4892

Qy       3807 TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 3866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4891 TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 4832

Qy       3867 ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCA 3926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4831 ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCA 4772

Qy       3927 GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACC 3986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4771 GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACC 4712

Qy       3987 GTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA 4046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4711 GTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA 4652

Qy       4047 ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT 4106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4651 ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT 4592

Qy       4107 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGT 4166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4591 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGT 4532

Qy       4167 GGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT 4226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4531 GGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT 4472

Qy       4227 AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC 4286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4471 AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC 4412

Qy       4287 TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4411 TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4352

Qy       4347 CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCAT 4406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4351 CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCAT 4292

Qy       4407 GTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACC 4466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4291 GTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACC 4232

Qy       4467 CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATC 4526
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4231 CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATC 4172

Qy       4527 TTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTACTAGGATCCGATATCA 4586
             ||||||||||||||||||||||||||||||||||||||||| |      ||     ||
Db       4171 TTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCCTGCA 4112

Qy       4587 TTTAAATCTAGGGATAACAGGGTAATACTAGTGTCGACCAACCAGATAAGTGAAATCTAG 4646
                                                 ||||||||||||||||||||||||||||||
Db       4111 ------------------------------GGTCGACCAACCAGATAAGTGAAATCTAG 4083

Qy       4647 TTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCC 4706
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4082 TTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCC 4023

Qy       4707 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAG 4766
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4022 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAG 3963

Qy       4767 TTCCCAACTATTTTCTACTCACTCAC 4792
             |||||||||||||| | | | | |||
Db       3962 TTCCCAACTATTTTGTCCGCCCACAC 3937
```

RESULT 6
AAD45060/c
ID AAD45060 standard; DNA; 5926 BP.
XX
AC AAD45060;
XX
DT 27-DEC-2002 (first entry)
XX
DE pEM delta transposon donor plasmid DNA.
XX
KW TnsC transposition regulatory protein; transposase; DNA sequencing;
KW genetic analysis; gene expression; ds.
XX
OS Unidentified.
XX
PN US6420524-B1.
XX
PD 16-JUL-2002.
XX
PF 20-FEB-1998; 98US-00027169.
XX
PR 20-FEB-1997; 97US-0037955P.
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Craig NL;
XX
DR WPI; 2002-634795/68.
XX
PT Mutant TnsC transposition regulatory protein for use in DNA sequencing,
PT for genetic analysis, and alteration of gene expression by insertion of a
PT given genetic sequence, comprises an alanine to valine substitution.
XX
PS Example 1; Col 71-76; 74pp; English.
XX
CC The invention relates to a mutant TnsC transposition regulatory protein.
CC This mutant protein activates a transposase in the absence of TnsD or
CC TnsE, which directs intermolecular transposition of a transposable
CC element and discriminates between immune and nonimmune targets, in a
CC manner which is characterised by reduced target site selectivity. It may
CC be used in DNA sequencing methods, for genetic analysis by insertional
CC mutagenesis and alteration of gene expression by insertion of a given
CC genetic sequence. The present DNA sequence is transposon donor plasmid
CC pEM delta, a pBR plasmid containing a kanamycin mTn7 element. This
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

Query Match 30.7%; Score 1477.2; DB 6; Length 5926;
Best Local Similarity 93.4%; Pred. No. 2.4e-267;
Matches 1594; Conservative 0; Mismatches 38; Indels 74; Gaps 2;

```
Qy       3087 TCCGTATGAGTGAGTAGAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA 3146
              ||   ||| | | | | ||||||||||||||||||||||||||||||||||||||||||||
Db       5568 TCTAAATGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA 5509

Qy       3147 TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAA 3206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5508 TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAA 5449

Qy       3207 AAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTG 3266
```

```
Db       5448 AAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTG 5389
Qy       3267 CCCGTCGTATTAAAGAGGGGCGTGGGGTCGACGCGGCCGCTAACTATAACGGTCCTAAGG 3326
Db       5388 CCCGTCGTATTAAAGAGGGGCGTGGGGTCGAC---------------------------- 5357
Qy       3327 TAGCGAGTTTAAACGATATCGGATCCGGCCGCCGCTGAGGTCTGCCTCGTGAAGAAGGTG 3386
Db       5356 --------------CTGCAGGGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTG 5312
Qy       3387 TTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAC 3446
Db       5311 TTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAC 5252
Qy       3447 GGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCA 3506
Db       5251 GGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCA 5192
Qy       3507 CGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAGAGTTC 3566
Db       5191 CGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTC 5132
Qy       3567 GATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAA 3626
Db       5131 GATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAA 5072
Qy       3627 CCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT 3686
Db       5071 CCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT 5012
Qy       3687 CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAA 3746
Db       5011 CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAA 4952
Qy       3747 CTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 3806
Db       4951 CTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 4892
Qy       3807 TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 3866
Db       4891 TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 4832
Qy       3867 ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCA 3926
Db       4831 ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCA 4772
Qy       3927 GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACC 3986
Db       4771 GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACC 4712
Qy       3987 GTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA 4046
Db       4711 GTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA 4652
Qy       4047 ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT 4106
Db       4651 ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT 4592
Qy       4107 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGT 4166
Db       4591 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGT 4532
Qy       4167 GGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT 4226
Db       4531 GGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT 4472
Qy       4227 AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC 4286
Db       4471 AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC 4412
Qy       4287 TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4346
```

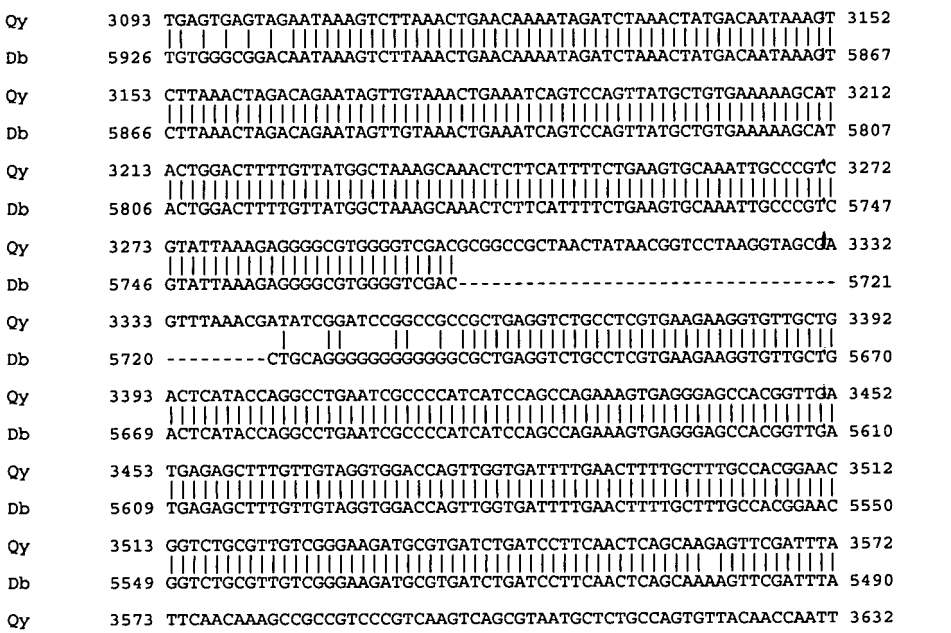
```
Db       4411 TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4352
Qy       4347 CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCAT 4406
Db       4351 CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCAT 4292
Qy       4407 GTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACC 4466
Db       4291 GTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACC 4232
Qy       4467 CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATC 4526
Db       4231 CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATC 4172
Qy       4527 TTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTACTAGGATCCGATATCA 4586
Db       4171 TTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCTGCA 4112
Qy       4587 TTTAAATCTAGGGATAACAGGGTAATACTAGTGTCGACCAACCAGATAAGTGAAATCTAG 4646
Db       4111 ------------------------------GGTCGACCAACCAGATAAGTGAAATCTAG 4083
Qy       4647 TTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCC 4706
Db       4082 TTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCC 4023
Qy       4707 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAG 4766
Db       4022 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAG 3963
Qy       4767 TTCCCAACTATTTTCTACTCACTCAC 4792
Db       3962 TTCCCAACTATTTTGTCCGCCCACAC 3937
```

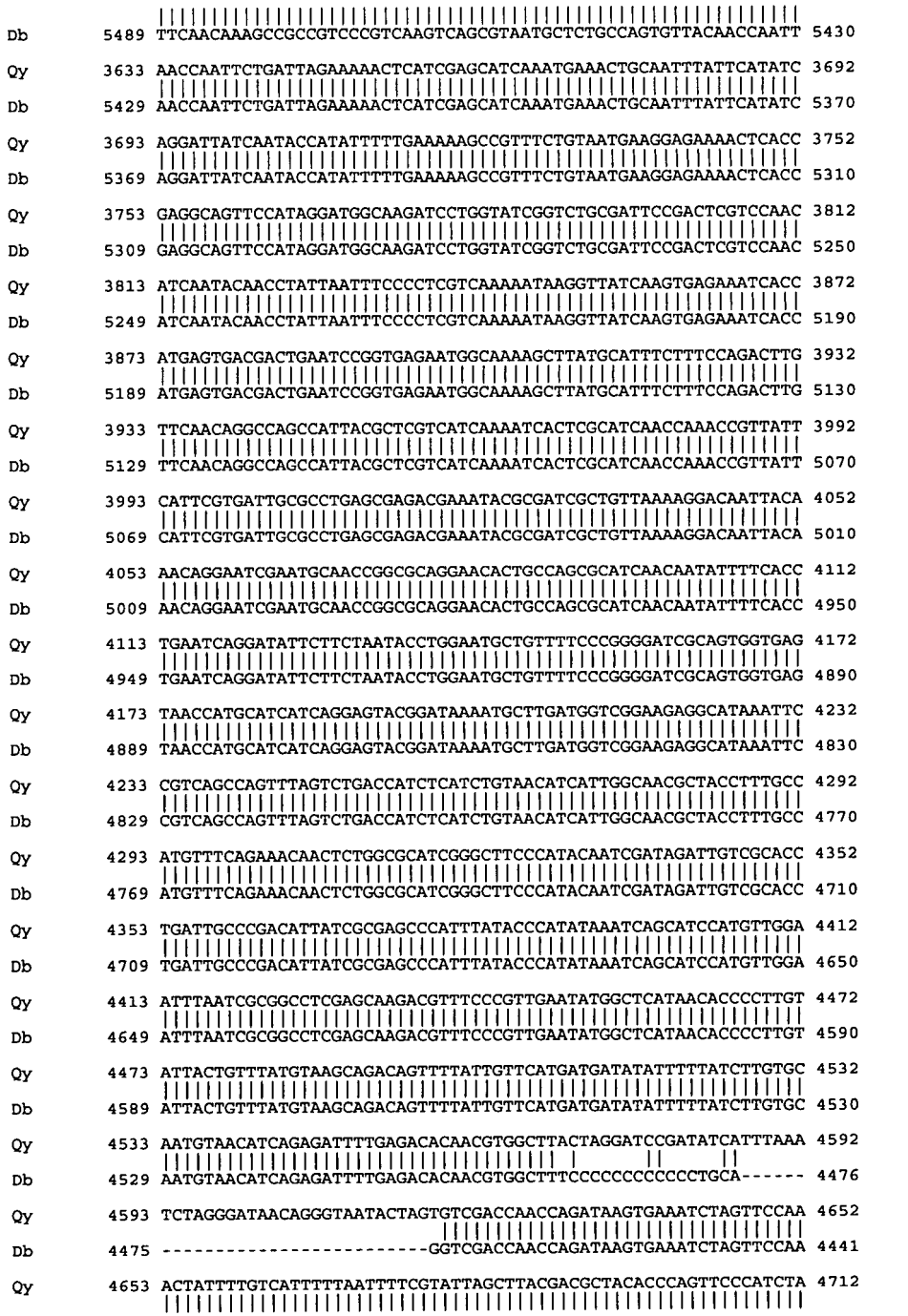
RESULT 7
AAV32977/c
ID AAV32977 standard; DNA; 5926 BP.
XX
AC AAV32977;
XX
DT 17-NOV-1998 (first entry)
XX
DE Tn7 donor plasmid sequence.
XX
KW Tn7; donor plasmid; transposon; ATP-utilising regulatory protein;
KW transposable element; DNA sequencing; genetic analysis;
KW insertional mutagenesis; ss.
XX
OS Synthetic.
XX
PN WO9837205-A1.
XX
PD 27-AUG-1998.
XX
PF 20-FEB-1998; 98WO-US003353.
XX
PR 20-FEB-1997; 97US-00037955.
XX
PA (CRAI/) CRAIG N L.
XX
PI Craig NL;
XX
DR WPI; 1998-467567/40.
XX
PT Transposon(s) encoding mutant ATP using proteins for insertion - which is
PT efficient and random, with reduced site specificity; for DNA sequencing
PT and altering gene expression.
XX
PS Disclosure; Fig 10B; 143pp; English.
XX
CC The present sequence represents a Tn7 donor plasmid pEM delta R.adj.1.
CC The plasmid carries a 1625 bp mini-Tn7 element: 199bp of Tn7L flank a

```
CC   kanamycin gene with SalI sites at the junctions. The backbone is pTRC99
CC   (Pharmacia). Transposon Tn7 encodes an ATP-utilising regulatory protein
CC   that contains a mutation that allows efficient and simple insertion of,
CC   and reduced target site specificity of, a transposable element derived
CC   from the transposon. The ATP-regulatory protein and compositions are used
CC   for the efficient, non-specific and simple insertion of a transposon or
CC   transposable element into a DNA segment. This is useful in DNA
CC   sequencing, for genetic analysis by insertional mutagenesis, or for
CC   alteration of gene expression by insertion of a desired sequence
XX
SQ   Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

  Query Match             30.7%;  Score 1476;  DB 2;  Length 5926;
  Best Local Similarity   93.6%;  Pred. No. 4.1e-267;
  Matches 1591;  Conservative    0;  Mismatches   35;  Indels   74;  Gaps    2;

Qy       3093 TGAGTGAGTAGAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT 3152
             || | | | | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5926 TGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT 5867

Qy       3153 CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCAT 3212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5866 CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCAT 5807

Qy       3213 ACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTC 3272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5806 ACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTC 5747

Qy       3273 GTATTAAAGAGGGGCGTGGGGTCGACGCGGCCGCTAACTATAACGGTCCTAAGGTAGCGA 3332
             ||||||||||||||||||||||||||
Db       5746 GTATTAAAGAGGGGCGTGGGGTCGAC---------------------------------- 5721

Qy       3333 GTTTAAACGATATCGGATCCGGCCGCCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTG 3392
                  |    ||    ||   | ||||||||||||||||||||||||||||||||||||
Db       5720 ---------CTGCAGGGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTG 5670

Qy       3393 ACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGA 3452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5669 ACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGA 5610

Qy       3453 TGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAAC 3512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5609 TGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAAC 5550

Qy       3513 GGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAGAGTTCGATTTA 3572
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       5549 GGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTA 5490

Qy       3573 TTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATT 3632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5489 TTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATT 5430

Qy       3633 AACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC 3692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5429 AACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC 5370

Qy       3693 AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACC 3752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5369 AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACC 5310

Qy       3753 GAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAAC 3812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5309 GAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAAC 5250

Qy       3813 ATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACC 3872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5249 ATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACC 5190

Qy       3873 ATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTG 3932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5189 ATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTG 5130

Qy       3933 TTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATT 3992
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 TTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATT 5070

Qy       3993 CATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACA 4052
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5069 CATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACA 5010

Qy       4053 AACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACC 4112
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5009 AACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACC 4950

Qy       4113 TGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAG 4172
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4949 TGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAG 4890

Qy       4173 TAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTC 4232
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4889 TAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTC 4830

Qy       4233 CGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCC 4292
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4829 CGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCC 4770

Qy       4293 ATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACC 4352
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4769 ATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACC 4710

Qy       4353 TGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGA 4412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4709 TGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGA 4650

Qy       4413 ATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGT 4472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4649 ATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGT 4590

Qy       4473 ATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGC 4532
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4589 ATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGC 4530

Qy       4533 AATGTAACATCAGAGATTTTGAGACACAACGTGGCTTACTAGGATCCGATATCATTTAAA 4592
             |||||||||||||||||||||||||||||||||||||||||||| |      ||    ||
Db       4529 AATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCCTGCA------ 4476

Qy       4593 TCTAGGGATAACAGGGTAATACTAGTGTCGACCAACCAGATAAGTGAAATCTAGTTCCAA 4652
                                           ||||||||||||||||||||||||||||||||
Db       4475 -------------------------GGTCGACCAACCAGATAAGTGAAATCTAGTTCCAA 4441

Qy       4653 ACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCCCATCTA 4712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4440 ACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCCCATCTA 4381

Qy       4713 TTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCA 4772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4380 TTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCA 4321

Qy       4773 ACTATTTTCTACTCACTCAC 4792
             |||||||| | | | | |||
Db       4320 ACTATTTTGTCCGCCCACAC 4301


RESULT 8
AAD45059/c
ID   AAD45059 standard; DNA; 5926 BP.
XX
AC   AAD45059;
XX
DT   27-DEC-2002  (first entry)
XX
DE   Transposon donor plasmid pEM delta R.adj to 1.
XX
KW   TnsC transposition regulatory protein; transposase; DNA sequencing;
KW   genetic analysis; gene expression; ds.
XX
```

OS Unidentified.
XX
PN US6420524-B1.
XX
PD 16-JUL-2002.
XX
PF 20-FEB-1998; 98US-00027169.
XX
PR 20-FEB-1997; 97US-0037955P.
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Craig NL;
XX
DR WPI; 2002-634795/68.
XX
PT Mutant TnsC transposition regulatory protein for use in DNA sequencing,
PT for genetic analysis, and alteration of gene expression by insertion of a
PT given genetic sequence, comprises an alanine to valine substitution.
XX
PS Example 4; Col 65-72; 74pp; English.
XX
CC The invention relates to a mutant TnsC transposition regulatory protein.
CC This mutant protein activates a transposase in the absence of TnsD or
CC TnsE, which directs intermolecular transposition of a transposable
CC element and discriminates between immune and nonimmune targets, in a
CC manner which is characterised by reduced target site selectivity. It may
CC be used in DNA sequencing methods, for genetic analysis by insertional
CC mutagenesis and alteration of gene expression by insertion of a given
CC genetic sequence. The present DNA sequence is transposon donor plasmid
CC pEM delta R.adj to 1, used in the exemplification of the invention. This
CC plasmid comprises a 1625 bp mini-Tn7 element:199 bp of Tn7R and 166 bp of
CC Tn7L flanking a Kan gene with SalI at the junctions
XX
SQ Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

Query Match 30.7%; Score 1476; DB 6; Length 5926;
Best Local Similarity 93.6%; Pred. No. 4.1e-267;
Matches 1591; Conservative 0; Mismatches 35; Indels 74; Gaps 2;





```
Db       4440 ACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCCCATCTA 4381

Qy       4713 TTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCA 4772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4380 TTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCA 4321

Qy       4773 ACTATTTTCTACTCACTCAC 4792
                |||||||| | | | | |||
Db       4320 ACTATTTTGTCCGCCCACAC 4301


RESULT 9
ADG46817/c
ID   ADG46817 standard; DNA; 5926 BP.
XX
AC   ADG46817;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Donor plasmid pEM delta R.adj to 1 DNA.
XX
KW   Transposon; ATP-utilising regulatory protein; genetic analysis;
KW   insertional mutagenesis; gene; plasmid ; cyclic; circular; ds.
XX
OS   Unidentified.
XX
PN   US2002188105-A1.
XX
PD   12-DEC-2002.
XX
PF   19-DEC-2001; 2001US-00024809.
XX
PR   20-FEB-1997;   97US-0037955P.
PR   20-FEB-1998;   98US-00027169.
XX
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Craig NL;
XX
DR   WPI; 2004-020185/02.
XX
PT   Transposon for use in e.g. DNA sequencing methods, comprises a mutation
PT   that allows efficient and simple insertion of and reduced target site
PT   specificity on the transposon.
XX
PS   Example 4; SEQ ID NO 3; 80pp; English.
XX
CC   The invention relates to a transposon encoding an ATP-utilising
CC   regulatory protein containing a mutation that allows efficient and simple
CC   insertion of and reduced target site specificity on the transposon. The
CC   invention is used in DNA sequencing methods, for genetic analysis by
CC   insertional mutagenesis, and alteration of gene expression. The invention
CC   is used to achieve efficient, simple, non-specific or random insertion
CC   into any given DNA segment. The present sequence is donor plasmid pEM
CC   delta R.adj to 1 DNA.
XX
SQ   Sequence 5926 BP; 1476 A; 1473 C; 1463 G; 1514 T; 0 U; 0 Other;

  Query Match             30.7%;  Score 1476;  DB 12;  Length 5926;
  Best Local Similarity   93.6%;  Pred. No. 4.1e-267;
  Matches 1591;  Conservative    0;  Mismatches   35;  Indels   74;  Gaps    2;

Qy       3093 TGAGTGAGTAGAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT 3152
             || | | | | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5926 TGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT 5867

Qy       3153 CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCAT 3212
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5866 CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAAAAGCAT 5807

Qy       3213 ACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTC 3272
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5806 ACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTGCCCGTC 5747

Qy       3273 GTATTAAAGAGGGGCGTGGGGTCGACGCGGCCGCTAACTATAACGGTCCTAAGGTAGCGA 3332
             |||||||||||||||||||||||||||
Db       5746 GTATTAAAGAGGGGCGTGGGGTCGAC---------------------------------- 5721

Qy       3333 GTTTAAACGATATCGGATCCGGCCGCCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTG 3392
                      |  ||     ||  | ||||||||||||||||||||||||||||||||||||
Db       5720 ---------CTGCAGGGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTGTTGCTG 5670

Qy       3393 ACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGA 3452
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5669 ACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCACGGTTGA 5610

Qy       3453 TGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAAC 3512
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5609 TGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCACGGAAC 5550

Qy       3513 GGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAGAGTTCGATTTA 3572
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       5549 GGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTCGATTTA 5490

Qy       3573 TTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATT 3632
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5489 TTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATT 5430

Qy       3633 AACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC 3692
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5429 AACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC 5370

Qy       3693 AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACC 3752
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5369 AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACC 5310

Qy       3753 GAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAAC 3812
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5309 GAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAAC 5250

Qy       3813 ATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACC 3872
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5249 ATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACC 5190

Qy       3873 ATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTG 3932
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5189 ATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTG 5130

Qy       3933 TTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATT 3992
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5129 TTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATT 5070

Qy       3993 CATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACA 4052
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5069 CATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACA 5010

Qy       4053 AACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACC 4112
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5009 AACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACC 4950

Qy       4113 TGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAG 4172
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4949 TGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAG 4890

Qy       4173 TAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTC 4232
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4889 TAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTC 4830

Qy       4233 CGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCC 4292
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4829 CGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCC 4770

Qy       4293 ATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACC 4352
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4769 ATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACC 4710
```

```
Qy       4353 TGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGA 4412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4709 TGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGA 4650

Qy       4413 ATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGT 4472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4649 ATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGT 4590

Qy       4473 ATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGC 4532
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4589 ATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGC 4530

Qy       4533 AATGTAACATCAGAGATTTTGAGACACAACGTGGCTTACTAGGATCCGATATCATTTAAA 4592
             ||||||||||||||||||||||||||||||||||| |       ||      ||
Db       4529 AATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCCTGCA------ 4476

Qy       4593 TCTAGGGATAACAGGGTAATACTAGTGTCGACCAACCAGATAAGTGAAATCTAGTTCCAA 4652
                                           ||||||||||||||||||||||||||||||||
Db       4475 -------------------------GGTCGACCAACCAGATAAGTGAAATCTAGTTCCAA 4441

Qy       4653 ACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCCCATCTA 4712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4440 ACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCCCATCTA 4381

Qy       4713 TTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCA 4772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4380 TTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAGTTCCCA 4321

Qy       4773 ACTATTTTCTACTCACTCAC 4792
             |||||||| | | | |||
Db       4320 ACTATTTTGTCCGCCCACAC 4301
```

RESULT 10
ADG46818/c
ID ADG46818 standard; DNA; 5926 BP.
XX
AC ADG46818;
XX
DT 11-MAR-2004 (first entry)
XX
DE pEM DNA (pBR plasmid containing a kanamycin mTn7 element).
XX
KW Transposon; ATP-utilising regulatory protein; genetic analysis;
KW insertional mutagenesis; gene; plasmid; cyclic; circular; ds.
XX
OS Unidentified.
XX
PN US2002188105-A1.
XX
PD 12-DEC-2002.
XX
PF 19-DEC-2001; 2001US-00024809.
XX
PR 20-FEB-1997; 97US-0037955P.
PR 20-FEB-1998; 98US-00027169.
XX
PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI Craig NL;
XX
DR WPI; 2004-020185/02.
XX
PT Transposon for use in e.g. DNA sequencing methods, comprises a mutation
PT that allows efficient and simple insertion of and reduced target site
PT specificity on the transposon.
XX
PS Example 2; SEQ ID NO 4; 80pp; English.
XX
CC The invention relates to a transposon encoding an ATP-utilising
CC regulatory protein containing a mutation that allows efficient and simple
CC insertion of and reduced target site specificity on the transposon. The
CC invention is used in DNA sequencing methods, for genetic analysis by
CC insertional mutagenesis, and alteration of gene expression. The invention
CC is used to achieve efficient, simple, non-specific or random insertion
CC into any given DNA segment. The present sequence is pEM DNA (pBR plasmid
CC containing a kanamycin mTn7 element) used in the exemplification of the
CC invention.
XX
SQ Sequence 5926 BP; 1475 A; 1471 C; 1464 G; 1516 T; 0 U; 0 Other;

Query Match 30.6%; Score 1474; DB 12; Length 5926;
Best Local Similarity 93.3%; Pred. No. 9.7e-267;
Matches 1592; Conservative 0; Mismatches 40; Indels 74; Gaps 2;

```
Qy       3087 TCCGTATGAGTGAGTAGAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA 3146
             ||   ||| | | | | ||||||||||||||||||||||||||||||||||||||||||||
Db       5568 TCTAAATGTGGGCGGACAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAA 5509

Qy       3147 TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAA 3206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5508 TAAAGTCTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGTGAAA 5449

Qy       3207 AAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTG 3266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5448 AAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCTGAAGTGCAAATTG 5389

Qy       3267 CCCGTCGTATTAAAGAGGGGCGTGGGGTCGACGCGGCCGCTAACTATAACGGTCCTAAGG 3326
             ||||||||||||||||||||||||||||||||||
Db       5388 CCCGTCGTATTAAAGAGGGGCGTGGGGTCGAC---------------------------- 5357

Qy       3327 TAGCGAGTTTAAACGATATCGGATCCGGCCGCCGCTGAGGTCTGCCTCGTGAAGAAGGTG 3386
                              |    ||    ||  | ||||||||||||||||||||||||||||
Db       5356 ---------------CTGCAGGGGGGGGGGGGCGCTGAGGTCTGCCTCGTGAAGAAGGTG 5312

Qy       3387 TTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAC 3446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5311 TTGCTGACTCATACCAGGCCTGAATCGCCCCATCATCCAGCCAGAAAGTGAGGGAGCCAC 5252

Qy       3447 GGTTGATGAGAGCTTTGTTGTAGGTGGACCAGTTGGTGATTTTGAACTTTTGCTTTGCCA 3506
             |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db       5251 GGTTGATGAGAGCTTTGTTGTAGGTGGACCAAATGGTGATTTTGAACTTTTGCTTTGCCA 5192

Qy       3507 CGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAGAGTTC 3566
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       5191 CGGAACGGTCTGCGTTGTCGGGAAGATGCGTGATCTGATCCTTCAACTCAGCAAAAGTTC 5132

Qy       3567 GATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAA 3626
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5131 GATTTATTCAACAAAGCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAA 5072

Qy       3627 CCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT 3686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5071 CCAATTAACCAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATT 5012

Qy       3687 CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAA 3746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5011 CATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAA 4952

Qy       3747 CTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 3806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4951 CTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG 4892

Qy       3807 TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 3866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4891 TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAA 4832

Qy       3867 ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCA 3926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4831 ATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCA 4772

Qy       3927 GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACC 3986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4771 GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACC 4712

Qy       3987 GTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA 4046
```

```
Db       4711 GTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACA 4652

Qy       4047 ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT 4106

Db       4651 ATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATT 4592

Qy       4107 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGT 4166

Db       4591 TTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGT 4532

Qy       4167 GGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT 4226

Db       4531 GGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT 4472

Qy       4227 AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC 4286

Db       4471 AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACC 4412

Qy       4287 TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4346

Db       4411 TTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT 4352

Qy       4347 CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCAT 4406

Db       4351 CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCAT 4292

Qy       4407 GTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACC 4466

Db       4291 GTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACC 4232

Qy       4467 CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATC 4526

Db       4231 CCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATC 4172

Qy       4527 TTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTACTAGGATCCGATATCA 4586

Db       4171 TTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCCCCCCCCCCCTGCA 4112

Qy       4587 TTTAAATCTAGGGATAACAGGGTAATACTAGTGTCGACCAACCAGATAAGTGAAATCTAG 4646

Db       4111 ------------------------------GGTCGACCAACCAGATAAGTGAAATCTAG 4083

Qy       4647 TTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCC 4706

Db       4082 TTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTCC 4023

Qy       4707 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAG 4766

Db       4022 CATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCACCCCTCCCAG 3963

Qy       4767 TTCCCAACTATTTTCTACTCACTCAC 4792

Db       3962 TTCCCAACTATTTTGTCCGCCCACAC 3937
```

```
RESULT 11
AAF59457
ID   AAF59457 standard; DNA; 4840 BP.
XX
AC   AAF59457;
XX
DT   03-MAY-2001  (first entry)
XX
DE   Npro-hGH expression plasmid pNPH1 nucleotide sequence.
XX
KW   Human; growth hormone; autoprotease; Npro; fusion protein;
KW   autoproteolytic activity; inclusion body; ds.
XX
OS   Homo sapiens.
OS   Classical swine fever virus.
XX
PN   WO200111057-A1.
XX
PD   15-FEB-2001.
XX
PF   07-AUG-2000; 2000WO-EP007643.
XX
PR   09-AUG-1999;   99AT-00001367.
XX
PA   (BIOC ) BIOCHEMIE GMBH.
XX
PI   Stempfer G,  Windisch J,  Knauseder F;
XX
DR   WPI; 2001-191545/19.
XX
PT   Producing heterologous polypeptide by culturing a bacterial cell
PT   expressing a fusion protein, comprising a polypeptide with
PT   autoproteolytic activity and a heterologous polypeptide, as an inclusion
PT   body.
XX
PS   Example 1; Page 15-17; 31pp; English.
XX
CC   The present invention describes a process for the recombinant production
CC   of desired heterologous polypeptides (HP). The method comprises culturing
CC   a bacterial cell expressing a fusion protein (FP), comprising a
CC   polypeptide with autoproteolytic activity of autoprotease Npro and a HP,
CC   in the form of cytoplasmic inclusion bodies. The inclusion bodies are
CC   isolated and treated so that HP is cleaved autoproteolytically by Npro
CC   activity from FP. The method is useful for producing a heterologous
CC   polypeptide with a clearly defined homogeneous N-terminus in a bacterial,
CC   preferably Escherichia coli cell. HP includes enzymes, agents with
CC   pharmaceutical activity such as interleukins, interferons such as
CC   leukocyte interferons, growth factors, in particular haemopoietic or
CC   wound-healing growth factors, such as granulocyte colony stimulating
CC   factor (G-CSF), erythropoietin, hormones such as human growth hormone
CC   (hGH), antibodies or vaccines. The present sequence represents the Npro-
CC   hGH expression plasmid pNPH1, where Npro is a classical swine fever virus
CC   autoprotease and hGH is human growth hormone, which is used in an example
CC   from the present invention
XX
SQ   Sequence 4840 BP; 1039 A; 1354 C; 1299 G; 1148 T; 0 U; 0 Other;
```

Query Match 30.1%; Score 1447.8; DB 4; Length 4840;
Best Local Similarity 99.9%; Pred. No. 7.8e-262;
Matches 1449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1328 GAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTT 1387

Db          1 GAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTT 60

Qy       1388 AAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCT 1447

Db         61 AAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCT 120

Qy       1448 CGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCT 1507

Db        121 CGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCT 180

Qy       1508 CTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCT 1567

Db        181 CTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCT 240

Qy       1568 ATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGG 1627

Db        241 ATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGG 300

Qy       1628 CCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGC 1687

Db        301 CCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGC 360

Qy       1688 GACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGC 1747

Db        361 GACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGC 420

Qy       1748 CACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCG 1807
```

```
Db        421 CACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCG 480

Qy       1808 CCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGG 1867

Db        481 CCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGG 540

Qy       1868 GGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGG 1927

Db        541 GGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGG 600

Qy       1928 CCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACC 1987

Db        601 CCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACC 660

Qy       1988 GATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTAT 2047

Db        661 GATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTAT 720

Qy       2048 CGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGC 2107

Db        721 CGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGC 780

Qy       2108 GCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTC 2167

Db        781 GCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTC 840

Qy       2168 GCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCAC 2227

Db        841 GCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCAC 900

Qy       2228 CAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTA 2287

Db        901 CAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTA 960

Qy       2288 CGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGC 2347

Db        961 CGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGC 1020

Qy       2348 TTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGA 2407

Db       1021 TTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGA 1080

Qy       2408 CCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGG 2467

Db       1081 CCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGG 1140

Qy       2468 ACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATG 2527

Db       1141 ACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATG 1200

Qy       2528 GATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAG 2587

Db       1201 GATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAG 1260

Qy       2588 CCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCA 2647

Db       1261 CCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCA 1320

Qy       2648 AGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAG 2707

Db       1321 AGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAG 1380

Qy       2708 AACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTT 2767

Db       1381 AACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTT 1440

Qy       2768 GGGTCCTGGGC 2778

Db       1441 GGGTCCTGGCC 1451
```

RESULT 12

```
AAQ29146/c
ID   AAQ29146 standard; DNA; 3474 BP.
XX
AC   AAQ29146;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   05-MAR-1993  (first entry)
XX
DE   pmTNF-MPH plasmid for expression of p362 from M. paratuberculosis.
XX
KW   Mycobacterium; Crohn's disease; Johne's disease; cattle; human; ss.
XX
OS   Mycobacterium avium subsp. paratuberculosis.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .208
FT                   /*tag=  a
FT                   /note= "lambda PL contg. EcoRI blunt-MboI blunt fragment
FT                   of pPL(lambda) "
FT   misc_feature    209. .436
FT                   /*tag=  b
FT                   /note= "synthetic DNA fragment"
FT   CDS             230. .307
FT                   /*tag=  c
FT                   /note= "encodes residues 2-25 of mature mouse TNF"
FT   misc_feature    308. .384
FT                   /*tag=  d
FT                   /note= "multiple cloning site contg. 6 His encoding
FT                   sequence at position 315-332"
FT   misc_feature    385. .436
FT                   /*tag=  e
FT                   /note= "HindIII fragment contg. E. coli trp terminator"
FT   misc_feature    437. .943
FT                   /*tag=  f
FT                   /note= "rrnBT1T2 contg. HindIII-SspI fragment from
FT                   pKK223"
FT   misc_feature    944. .3474
FT                   /*tag=  g
FT                   /note= "DraI-EcoRI blunt fragment of pAT153 contg.
FT                   tetracycline resistance gene and origin of replication"
XX
PN   WO9216628-A1.
XX
PD   01-OCT-1992.
XX
PF   24-MAR-1992;   92WO-EP000661.
XX
PR   25-MAR-1991;   91EP-00400798.
XX
PA   (INNO-) INNOGENETICS NV SA.
XX
PI   Cocito C,  Coene M,  De Kesel M,  Gilot P;
XX
DR   WPI; 1992-349213/42.
XX
PT   Polypeptide(s) derived from mycobacterium para-tuberculosis - are
PT   immunogens, useful as vaccines and for diagnosing Crohns disease.
XX
PS   Disclosure; Fig 9; 105pp; English.
XX
CC   The DNA sequence is that of the expression vector pmTNF-MPH which
CC   contains the tetracycline resistance gene and origin of replication of
CC   pAT153, the lambda PL promoter up to the MboII site in the N gene 5'
CC   untranslated region, followed by a synthetic ribosome binding site and
CC   the information encoding the first 25 amino acids of mouse TNF. This
CC   sequence is followed by a synthetic polylinker encoding six consecutive
CC   histidines followed by several proteolytic sites each accessible via a
CC   different restriction enzyme which is unique for the plasmid. Downstream
CC   from the polylinker, several transcription terminators are present,
CC   including the E. coli trp terminator. The DNA encoding a fragment of the
CC   34 kD p362 protein from Mycobacterium paratuberculosis was subcloned into
CC   this expression vector to express the fusion protein mTNF-H6-p362. The
```

```
CC   purified fusion protein may be used to assay for M. paratuberculosis
CC   infections, e.g. Johne's disease in cattle or other animals and possibly
CC   Crohn's disease in humans. See also AAQ29145-7. (Updated on 25-MAR-2003
CC   to correct PN field.) (Updated on 25-MAR-2003 to correct PI field.)
CC   (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 3474 BP; 844 A; 933 C; 978 G; 717 T; 0 U; 2 Other;

  Query Match           30.1%;  Score 1446.8;  DB 2;  Length 3474;
  Best Local Similarity  99.9%;  Pred. No. 1.2e-261;
  Matches 1448;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1329 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 3415

Qy       1389 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 1448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 3355

Qy       1449 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 3295

Qy       1509 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 3235

Qy       1569 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 3175

Qy       1629 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 3115

Qy       1689 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 3055

Qy       1749 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 1808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 2995

Qy       1809 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 1868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 2935

Qy       1869 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 1928
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2934 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 2875

Qy       1929 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 1988
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2874 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 2815

Qy       1989 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2814 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2755

Qy       2049 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2754 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2695

Qy       2109 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2168
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2694 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2635

Qy       2169 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2634 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2575

Qy       2229 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2574 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2515

Qy       2289 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2455

Qy       2349 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2395

Qy       2409 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGA 2468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2394 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGA 2335

Qy       2469 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2275

Qy       2529 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2588
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2215

Qy       2589 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2155

Qy       2649 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2095

Qy       2709 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2094 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2035

Qy       2769 GGTCCTGGGC 2778
             |||||||| |
Db       2034 GGTCCTGGCC 2025


RESULT 13
AAQ51546/c
ID   AAQ51546 standard; DNA; 3474 BP.
XX
AC   AAQ51546;
XX
DT   25-MAR-2003  (revised)
DT   24-MAY-1994  (first entry)
XX
DE   Expression plasmid pmTNF-MPH.
XX
KW   Macrophage; induced; lipo-polyasaccharide; antitumour; antiinflammatory;
KW   trypanocidal agent; antibody; cell prolification; activation;
KW   cytotoxicity; ds.
XX
OS   Synthetic.
XX
PN   WO9322437-A1.
XX
PD   11-NOV-1993.
XX
PF   28-APR-1993;   93WO-EP001022.
XX
PR   30-APR-1992;   92EP-00401231.
XX
PA   (INNO-) INNOGENETICS NV SA.
XX
PI   Fransen L,  Devos K,  Van De Voorde A,  Van Heuverswyn H;
XX
DR   WPI; 1993-368796/46.
XX
PT   New polypeptide induced in macrophage(s) by lipo-polysaccharide - useful
PT   e.g. as antitumour, antiinflammatory or trypanocidal agent, also related
```

```
PT   nucleic acid, antibodies, anti-sense cpds. etc.
XX
PS   Disclosure; Fig 8b; 108pp; English.
XX
CC   The plasmid is used to clone and express. The polypeptide induced in
CC   macrophages by lipopolysaccharide stimulates cell prolification (esp.
CC   when costimulated with IL-4) promote activation, cytotoxicity, and
CC   mobilisation of LAK cells; promote recruitment of suppressive peritoneal
CC   exudate cells; promote generation of immunocompetent lymph node cells
CC   (LNC) and have trypanocidal and trypanolytic activity. The human and
CC   murine sequences are given in (AAQ51543-45), peptide fragments able to
CC   generate antibodies are given in (AAR51951-61) (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 3474 BP; 845 A; 933 C; 978 G; 718 T; 0 U; 0 Other;

  Query Match             30.1%;  Score 1446.8;  DB 2;  Length 3474;
  Best Local Similarity   99.9%;  Pred. No. 1.2e-261;
  Matches 1448;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1329 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3474 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 3415

Qy       1389 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 1448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3414 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 3355

Qy       1449 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3354 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 3295

Qy       1509 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3294 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 3235

Qy       1569 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3234 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 3175

Qy       1629 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3174 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 3115

Qy       1689 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3114 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 3055

Qy       1749 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 1808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3054 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 2995

Qy       1809 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 1868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2994 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 2935

Qy       1869 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 1928
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2934 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 2875

Qy       1929 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 1988
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2874 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 2815

Qy       1989 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2814 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2755

Qy       2049 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2754 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2695

Qy       2109 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2168
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2694 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2635

Qy       2169 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2634 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2575

Qy       2229 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2574 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2515

Qy       2289 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2514 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2455

Qy       2349 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2454 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2395

Qy       2409 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGA 2468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       2394 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGA 2335

Qy       2469 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2334 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2275

Qy       2529 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2588
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2274 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2215

Qy       2589 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2214 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2155

Qy       2649 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2154 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2095

Qy       2709 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2094 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2035

Qy       2769 GGTCCTGGGC 2778
             |||||||| |
Db       2034 GGTCCTGGCC 2025


RESULT 14
ADL72227/c
ID   ADL72227 standard; DNA; 3801 BP.
XX
AC   ADL72227;
XX
DT   01-JUL-2004  (first entry)
XX
DE   DNA sequence of plasmid construct, pACSE.
XX
KW   Evolutionary potential; mutant resistance gene; anti-pathogenic drug;
KW   pathogen; ds.
XX
OS   Synthetic.
XX
PN   US6720142-B1.
XX
PD   13-APR-2004.
XX
PF   18-AUG-2000; 2000US-00640882.
XX
PR   19-AUG-1999;   99US-0149813P.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
```

```
PI   Hall BG;
XX
DR   WPI; 2004-313650/29.
XX
PT   Predicting the evolutionary potential of a mutant resistance gene, useful
PT   for screening for anti-pathogenic drugs comprises determining if a mutant
PT   resistance gene is likely to evolve through two or more independent
PT   mutation events.
XX
PS   Disclosure; SEQ ID NO 1; 47pp; English.
XX
CC   The present invention relates to a method of predicting the evolutionary
CC   potential of a mutant resistance gene. The method comprises providing a
CC   host cell which includes a mutant resistance gene and determining whether
CC   the mutant resistance gene is likely to evolve through two or more
CC   independent mutation events. Also disclosed are mutant resistance genes
CC   and the encoded mutant resistance-conferring polypeptides, and a method
CC   of screening a drug for anti-pathogenic activity against a pathogen and
CC   assessing the potential longevity of a candidate anti-pathogenic drug.
CC   The mutant resistance gene includes two or more nucleic acid
CC   modifications, which affects the expression levels of the encoded
CC   polypeptide. The mutated resistance gene is inserted into a host cell
CC   using a plasmid and treating the host cell to incorporate the plasmid
CC   into the host cell. The method is useful for predicting the evolutionary
CC   potential of a mutant resistance gene. The mutant resistance genes are
CC   useful for screening for a drug for anti-pathogenic activity against a
CC   pathogen. The present sequence represents a plasmid construct used in the
CC   method of the invention.
XX
SQ   Sequence 3801 BP; 939 A; 978 C; 1028 G; 856 T; 0 U; 0 Other;

  Query Match             30.1%;  Score 1446.8;  DB 12;  Length 3801;
  Best Local Similarity   99.9%;  Pred. No. 1.2e-261;
  Matches 1448;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1329 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2308 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 2249

Qy       1389 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 1448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2248 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 2189

Qy       1449 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2188 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 2129

Qy       1509 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2128 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 2069

Qy       1569 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2068 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 2009

Qy       1629 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2008 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1949

Qy       1689 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1948 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1889

Qy       1749 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 1808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1888 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 1829

Qy       1809 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 1868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1828 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 1769

Qy       1869 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 1928
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1768 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 1709

Qy       1929 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 1988
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1708 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 1649

Qy       1989 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1648 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 1589

Qy       2049 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1588 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 1529

Qy       2109 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2168
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1528 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 1469

Qy       2169 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1468 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 1409

Qy       2229 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1408 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 1349

Qy       2289 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1348 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 1289

Qy       2349 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1288 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 1229

Qy       2409 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGA 2468
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       1228 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGA 1169

Qy       2469 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1168 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 1109

Qy       2529 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2588
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1108 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 1049

Qy       2589 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1048 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 989

Qy       2649 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 929

Qy       2709 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        928 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 869

Qy       2769 GGTCCTGGGC 2778
             |||||||| |
Db        868 GGTCCTGGCC 859


RESULT 15
ABZ70623
ID   ABZ70623 standard; DNA; 4245 BP.
XX
AC   ABZ70623;
XX
DT   23-MAY-2003  (first entry)
XX
DE   Plasmid pACYC184.
XX
```

KW Vector; pACYC184; origin of replication; p15A; ds.
XX
OS Escherichia coli.
XX
PN WO2003008556-A2.
XX
PD 30-JAN-2003.
XX
PF 18-JUL-2002; 2002WO-US023089.
XX
PR 18-JUL-2001; 2001US-0306344P.
XX
PA (EXPR-) EXPRESSIVE CONSTRUCTS INC.
XX
PI Sanders MC, Hamilton M;
XX
DR WPI; 2003-239326/23.
XX
PT New p15A origin of replication, useful in expressing and/or co-expressing
PT proteins, in purifying a cloned nucleic acid, and in converting a low
PT copy number plasmid into a more efficient high copy number plasmid.
XX
PS Disclosure; Page 48-49; 53pp; English.
XX
CC The present sequence is the DNA sequence of plasmid pACYC184. The plasmid
CC includes an origin of replication sequence, denoted p15A ori, positioned
CC within an EcoRI-HindIII fragment corresponding to nucleotides 1-1523 of
CC the present sequence. The present invention is based on the finding that
CC key mutations in the p15A ori sequence operatively linked within a
CC plasmid result in the plasmid being converted from a low copy number to a
CC high copy number plasmid useful for more efficient cloning, DNA and RNA
CC purification, protein expression and co-expression. The mutated
CC (enhanced) p15A ori can be isolated and inserted into low copy plasmids
CC other than pACYC184 to convert them to high copy number plasmids.
CC Preferred mutations to p15A ori were identified following random
CC mutagenesis of the native p15A locus of pACYC184 by PCR, use of an
CC amplified 1.5 kb product to construct a plasmid library, transformation
CC of Escherichia coli BL21(DE3) cells and screening of transformed colonies
CC for an operationally enhanced p15A ori. Plasmid pECI-073 was obtained,
CC which has a copy number in E. coli BL21(DE3) about 4.47-fold higher than
CC that of pACYC184. Sequencing of the EcoRI-HindIII fragment of pECI-073
CC revealed 10 base substitutions (see ABZ70621) compared with the native
CC sequence
XX
SQ Sequence 4245 BP; 956 A; 1167 C; 1094 G; 1028 T; 0 U; 0 Other;

Query Match 30.1%; Score 1446.8; DB 8; Length 4245;
Best Local Similarity 99.9%; Pred. No. 1.2e-261;
Matches 1448; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1329 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1494 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1553

Qy      1389 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 1448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1554 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 1613

Qy      1449 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1614 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 1673

Qy      1509 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1674 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 1733

Qy      1569 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1734 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 1793

Qy      1629 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1794 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1853

Qy      1689 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1854 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1913

Qy      1749 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 1808
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1914 ACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGC 1973

Qy      1809 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 1868
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1974 CACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGG 2033

Qy      1869 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 1928
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2034 GGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGC 2093

Qy      1929 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 1988
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2094 CTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCG 2153

Qy      1989 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2048
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2154 ATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATC 2213

Qy      2049 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2108
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2214 GTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCG 2273

Qy      2109 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2168
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2274 CTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCG 2333

Qy      2169 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2228
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2334 CTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACC 2393

Qy      2229 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2288
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2394 AAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTAC 2453

Qy      2289 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2454 GTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCT 2513

Qy      2349 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2514 TCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGAC 2573

Qy      2409 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGA 2468
             |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db      2574 CATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCACTGGA 2633

Qy      2469 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2634 CCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGG 2693

Qy      2529 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2588
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2694 ATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGC 2753

Qy      2589 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2754 CGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAA 2813

Qy      2649 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2708
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2814 GAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGA 2873

Qy      2709 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2768
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2874 ACATATCCATCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTG 2933
```

```
Qy       2769 GGTCCTGGGC 2778
              ||||||||| |
Db       2934 GGTCCTGGCC 2943
```

Search completed: January 17, 2006, 19:35:29
Job time : 3011.54 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 17, 2006, 18:09:28 ; Search time 20375.3 Seconds
(without alignments)
11054.239 Million cell updates/sec

Title: US-10-511-327-7
Perfect score: 4814
Sequence: 1 ggtaccctgtgaatgcgcaa.....taagatgcttttctgtgact 4814

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	920.4	19.1	933	9	BZ430219	BZ430219 BONDX17TF
	2	918	19.1	918	9	BZ430224	BZ430224 BONDX17TR
c	3	801	16.6	881	9	AZ682163	AZ682163 ENTIW91TR
	4	719.2	14.9	988	11	CNS06X6V	AL419357 T7 end of
	5	704.8	14.6	889	6	CF577494	CF577494 AGENCOURT
c	6	670.2	13.9	675	9	BH600610	BH600610 BOHKQ87TF
	7	658	13.7	663	10	CG411076	CG411076 RM391 Lxx
	8	648	13.5	648	1	AJ762384	AJ762384 AJ762384
	9	648	13.5	648	1	AJ762542	AJ762542 AJ762542
	10	646	13.4	659	2	BF632491	BF632491 NF027F11D
	11	645	13.4	645	5	BQ752029	BQ752029 EST632592
c	12	643	13.4	681	10	CG804998	CG804998 1118056B1
	13	641	13.3	675	3	BJ684207	BJ684207 BJ684207
	14	641	13.3	676	3	BJ683711	BJ683711 BJ683711
	15	637.4	13.2	683	3	BJ683533	BJ683533 BJ683533
	16	636.4	13.2	675	3	BJ683290	BJ683290 BJ683290
	17	636	13.2	675	3	BJ683026	BJ683026 BJ683026
c	18	635.4	13.2	945	9	BH146743	BH146743 ENTPJ23TF
	19	635	13.2	651	3	BJ684301	BJ684301 BJ684301
	20	635	13.2	670	3	BJ683987	BJ683987 BJ683987
c	21	628.4	13.1	947	9	AZ541608	AZ541608 ENTFU74TF
	22	627.6	13.0	643	3	BJ682782	BJ682782 BJ682782
c	23	620.8	12.9	647	9	BH241665	BH241665 ATYQB72TR
	24	617.8	12.8	650	3	BJ683212	BJ683212 BJ683212
	25	616	12.8	645	3	BJ682197	BJ682197 BJ682197
	26	613.2	12.7	658	3	BJ684166	BJ684166 BJ684166
	27	612.2	12.7	785	2	BE572157	BE572157 601330463
c	28	604.4	12.6	618	9	CC943909	CC943909 BOIHE17TR
c	29	604	12.5	979	9	BH146752	BH146752 ENTPJ23TR
c	30	591.8	12.3	599	9	CC964255	CC964255 BOIEB66TF
	31	587.2	12.2	710	10	AG011497	AG011497 Homo sapi
	32	584	12.1	638	1	AJ762641	AJ762641 AJ762641
c	33	582.6	12.1	1059	1	AL575254	AL575254 AL575254
	34	581.8	12.1	585	5	BU668685	BU668685 MC01033H0
c	35	572	11.9	572	9	CC948132	CC948132 BOIEC75TF
	36	566.2	11.8	579	5	BU719372	BU719372 SJM2CTE08
c	37	565.2	11.7	570	9	BH241973	BH241973 ATYQC33TF
c	38	561.8	11.7	580	9	CC967108	CC967108 BOIHM43TF
	39	560	11.6	574	7	CJ326964	CJ326964 CJ326964
c	40	556.6	11.6	745	9	BZ054578	BZ054578 jnr34a02.
c	41	554.6	11.5	716	9	BZ050418	BZ050418 jnr41b09.
	42	546.4	11.4	561	3	BJ682307	BJ682307 BJ682307
c	43	544.6	11.3	716	9	BZ050188	BZ050188 jnr43b08.
c	44	543.6	11.3	715	9	BZ054871	BZ054871 jnr32h05.
c	45	543.6	11.3	721	9	BZ054951	BZ054951 jnr32h05.

ALIGNMENTS

RESULT 1
BZ430219/c
LOCUS BZ430219 933 bp DNA linear GSS 13-DEC-2002
DEFINITION BONDX17TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONDX17, genomic survey sequence.
ACCESSION BZ430219
VERSION BZ430219.1 GI:26675100
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 933)
AUTHORS Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S., Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
TITLE Whole genome shotgun sequencing of Brassica oleracea and its application to gene discovery and annotation in Arabidopsis
JOURNAL Genome Res. 15 (4), 487-495 (2005)
PUBMED 15805490
COMMENT Other_GSSs: BONDX17TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .933
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BONDX17"
/clone_lib="BO_1.6_2_KB_tot"
/note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

Query Match 19.1%; Score 920.4; DB 9; Length 933;
Best Local Similarity 99.8%; Pred. No. 2.1e-246;
Matches 932; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy       1786 CCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTA 1845
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        933 CCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTA 874

Qy       1846 TGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCC 1905
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        873 TGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCC 814

Qy       1906 TTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGT 1965
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        813 TTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGT 754

Qy       1966 CGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCC 2025
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        753 CGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCC 694

Qy       2026 GGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAAC 2085
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        693 GGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAAC 634

Qy       2086 TCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGA 2145
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        633 TCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGA 574

Qy       2146 GCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAG 2205
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        573 GCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAG 514

Qy       2206 CCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCA 2265
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        513 CCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCA 454

Qy       2266 TGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCT 2325
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        453 TGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCT 394

Qy       2326 TCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGC 2385
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        393 TCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGC 334

Qy       2386 TGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTA 2445
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 TGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTA 274

Qy       2446 CCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGA 2505
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 CCAGCCTAACTTCGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGA 214

Qy       2506 GCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCG 2565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 GCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCG 154

Qy       2566 CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCT 2625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 CGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCT 94

Qy       2626 CGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAA 2685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db          93 CGCTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACAGTGAA 34

Qy       2686 TGCGCAAACCAACCCTTGGCAGAACATATCCATC 2719
             ||||||||||||||| ||||||||||||||||||
Db          33 TGCGCAAACCAACCC-TGGCAGAACATATCCATC 1
```

RESULT 2
BZ430224

```
LOCUS       BZ430224                 918 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONDX17TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONDX17,
            genomic survey sequence.
ACCESSION   BZ430224
VERSION     BZ430224.1  GI:26675106
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
            Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea and its
            application to gene discovery and annotation in Arabidopsis
  JOURNAL   Genome Res. 15 (4), 487-495 (2005)
   PUBMED   15805490
COMMENT     Other_GSSs: BONDX17TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONDX17"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN
```

Query Match 19.1%; Score 918; DB 9; Length 918;
Best Local Similarity 100.0%; Pred. No. 9.6e-246;
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1358 TAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTAT 1417
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTAT 60

Qy       1418 GAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCAT 1477
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCAT 120

Qy       1478 AGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCAT 1537
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCAT 180

Qy       1538 CGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACC 1597
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACC 240

Qy       1598 CGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACT 1657
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACT 300

Qy       1658 TGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGC 1717
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGC 360

Qy       1718 CGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGC 1777
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGC 420

Qy       1778 CGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGG 1837
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGG 480

Qy       1838 CGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGC 1897
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        481 CGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGC 540

Qy       1898 ACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAAT 1957
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAAT 600

Qy       1958 GCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAG 2017
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAG 660

Qy       2018 CTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTAT 2077
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTAT 720

Qy       2078 CATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTT 2137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTT 780

Qy       2138 TCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCT 2197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCT 840

Qy       2198 CGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTAT 2257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTAT 900

Qy       2258 CGCCGGCATGGCGGCCGA 2275
              ||||||||||||||||||
Db        901 CGCCGGCATGGCGGCCGA 918
```

```
RESULT 3
AZ682163/c
LOCUS       AZ682163                881 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  ENTIW91TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   AZ682163
VERSION     AZ682163.1  GI:11819309
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 816.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
ORIGIN
```

```
  Query Match             16.6%;  Score 801;  DB 9;  Length 881;
  Best Local Similarity   97.2%;  Pred. No. 7.2e-213;
  Matches  837;  Conservative    0;  Mismatches   20;  Indels    4;  Gaps    2;

Qy       1109 TGTGGAACACCTACATCTGTATTAACGAAGCATTATTGAAGCATTTATCAGGGTTATTGT 1168
              ||| ||| ||  | |    | |     | |||||  ||||||||||||||||||||||
Db        875 TGTTGAATACTCATACTCTTCCTTTTTCAATATTATGAAAGCATTTATCAGGGTTATTGT 816

Qy       1169 CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGC 1228
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        815 CTCATGAGCGGATACATATTTGAATGTA-TTAGAAAAATAAACAAATAGGGGTTCCGCGC 757

Qy       1229 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 1288
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 ACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC 697

Qy       1289 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAG---AATTCTCATGTTTGACA 1345
              ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        696 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTTCAAGAATAATTCTCATGTTTGACA 637

Qy       1346 GCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCA 1405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCA 577

Qy       1406 GGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGA 1465
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        576 GGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGA 517

Qy       1466 TGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCA 1525
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 TGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCA 457

Qy       1526 TTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATT 1585
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 TTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATT 397

Qy       1586 TCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCT 1645
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        396 TCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCT 337

Qy       1646 CGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTG 1705
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        336 CGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTG 277

Qy       1706 GATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGG 1765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 GATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGG 217

Qy       1766 CGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAG 1825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 CGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAG 157

Qy       1826 CGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCAT 1885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        156 CGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCAT 97

Qy       1886 CTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGG 1945
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         96 CTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGG 37

Qy       1946 CTGCTTCCTAATGCAGGAGTC 1966
              |||||||||||||||||||||
Db         36 CTGCTTCCTAATGCAGGAGTC 16
```

```
RESULT 4
CNS06X6V
LOCUS       CNS06X6V                988 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T7 end of clone AX0AA038H01 of library AX0AA from strain CBS 7064
            of Pichia farinosa, genomic survey sequence.
ACCESSION   AL419357
VERSION     AL419357.1  GI:12202535
KEYWORDS    GSS.
SOURCE      Pichia farinosa
  ORGANISM  Pichia farinosa
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   2  (bases 1 to 988)
  AUTHORS   de Montigny,J., Spehner,C., Souciet,J., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Potier,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
            sorbitophila
  JOURNAL   FEBS Lett. 487 (1), 87-90 (2000)
   PUBMED   11152890
REFERENCE   3  (bases 1 to 988)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .988
                     /organism="Pichia farinosa"
                     /mol_type="genomic DNA"
                     /strain="CBS 7064"
                     /db_xref="taxon:4920"
                     /clone="AX0AA038H01"
                     /clone_lib="AX0AA"
                     /note="end : T7"
ORIGIN

  Query Match             14.9%;  Score 719.2;  DB 11;  Length 988;
  Best Local Similarity   99.0%;  Pred. No. 7.3e-190;
  Matches  721;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       3636 CAATTCTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGG 3695
              || | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 CAGTACTGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGG 320

Qy       3696 ATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAG 3755
              |||||||||||  |||||||:||||||||||||||||||||||||||||||||||||||
Db        321 ATTATCAATACTTTATTTTTGAMAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAG 380

Qy       3756 GCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATC 3815
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATC 440

Qy       3816 AATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATG 3875
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 AATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATG 500

Qy       3876 AGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTC 3935
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AGTGACGACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTC 560

Qy       3936 AACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCAT 3995
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 AACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCAT 620

Qy       3996 TCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAAC 4055
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 TCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAAC 680

Qy       4056 AGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGA 4115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 AGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGA 740

Qy       4116 ATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAA 4175
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        741 ATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTGCCGGGGATCGCAGTGGTGAGTAA 800

Qy       4176 CCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGT 4235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        801 CCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGT 860

Qy       4236 CAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATG 4295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 CAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATG 920

Qy       4296 TTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGA 4355
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        921 TTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGA 980

Qy       4356 TTGCCCGA 4363
              :|||||||
Db        981 WTGCCCGA 988


RESULT 5
CF577494
LOCUS       CF577494                889 bp    mRNA    linear   EST 24-SEP-2003
DEFINITION  AGENCOURT_13211379_updated NIH_MGC_137 Mus musculus cDNA clone
            IMAGE:6436303 5', mRNA sequence.
ACCESSION   CF577494
VERSION     CF577494.1  GI:35190021
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)
             cDNA Library Preparation: Catherine Lee, Endocrine Pancreas
            Consortium
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: IRBD67  row: a  column: 08
```

```
          High quality sequence start: 16
          High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6436303"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_137"
                     /note="Organ: pancreas; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; Library consists of a pool of clones
                     rearrayed from the following libraries: Melton normalized
                     mixed mouse pancreas 1 N1-MMS1, Amplified Melton mouse
                     islets 1 MIS1-A, and Kaestner ngn3 wt. Clones rearrayed in
                     the laboratory of K. Kaestner (University of
                     Pennsylvania). Note: this is a NIH_MGC Library."
ORIGIN

  Query Match           14.6%;  Score 704.8;  DB 6;  Length 889;
  Best Local Similarity  97.4%;  Pred. No. 7.8e-186;
  Matches  738;  Conservative    0;  Mismatches   17;  Indels    3;  Gaps    2;

Qy       3582 GCCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTC 3641
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 GGCGCCGTCCCGTCAAGTCAGCGTAATGCTCTGCCAGTGTTACAACCAATTAACCAATTC 179

Qy       3642 TGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC 3701
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 TGATTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATC 239

Qy       3702 AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTT 3761
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 AATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTT 299

Qy       3762 CCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACA 3821
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 CCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACA 359

Qy       3822 ACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGAC 3881
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 ACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGAC 419

Qy       3882 GACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGG 3941
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        420 GACTGAATCCGGTGAGAATGGCAAAAACTTATGCATTTCTTTCCAGACTTGTTCAACAGG 479

Qy       3942 CCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGA 4001
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 CCAACCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGA 539

Qy       4002 TTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAAT 4061
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 TTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAAT 599

Qy       4062 CGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGG 4121
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 CGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGG 659

Qy       4122 ATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGC 4181
             |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        660 ATATTCTTCTAATACCTGGAATGCTGTTTTACCGGGGATCGCAGTGGTGAGTAACCATGC 719

Qy       4182 ATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCA 4241
             |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        720 ATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAAGCATAAATTCCGTCAGCCA 779

Qy       4242 GTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATG-TTTCA 4300
             |||||||||||||||||||||||||||||||| |  |||||||||||  |||| |||||
Db        780 GTTTAGTCTGACCATCTCATCTGTAACATCATTTGGCACGCTACCTTTTGCATGTTTTCA 839

Qy       4301 GAAACAACTCTGG--CGCATCGGGCTTCCCATACAATC 4336
              ||||||||||||       |||||||||| |||||||
Db        840 AAAACAACTCTGGGCGCATCCGGGCTTCCCTTACAATC 877


RESULT 6
BH600610/c
LOCUS       BH600610                675 bp    DNA     linear   GSS 15-DEC-2001
DEFINITION  BOHKQ87TF BOHK Brassica oleracea genomic clone BOHKQ87, genomic
            survey sequence.
ACCESSION   BH600610
VERSION     BH600610.1  GI:17853056
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
            Utterback,T.R., Wortman,J.R., White,O.R. and Town,C.D.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea and its
            application to gene discovery and annotation in Arabidopsis
  JOURNAL   Genome Res. 15 (4), 487-495 (2005)
   PUBMED   15805490
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHKQ87"
                     /clone_lib="BOHK"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match           13.9%;  Score 670.2;  DB 9;  Length 675;
  Best Local Similarity  99.6%;  Pred. No. 3.8e-176;
  Matches  672;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2047 TCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAG 2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         675 TCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAG 616

Qy       2107 CGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGT 2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         615 CGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGT 556

Qy       2167 CGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCA 2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         555 CGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCA 496

Qy       2227 CCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCT 2286
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         495 CCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCT 436

Qy       2287 ACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCG 2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         435 ACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCG 376

Qy       2347 CTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACG 2406
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         375 CTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACG 316

Qy       2407 ACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTG 2466
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 ACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTG 256

Qy       2467 GACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCAT 2526
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 GACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCAT 196

Qy       2527 GGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGA 2586
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 GGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGA 136

Qy       2587 GCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCC 2646
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 GCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCC 76

Qy       2647 AAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCA 2706
                ||||||||||||||||||||||||||||||||||||||||||||||||||  || ||||||
Db         75 AAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCGGCCATTGGCA 16

Qy       2707 GAACATATCCATCGC 2721
                ||||||||||||||
Db         15 GAACATATCCATCGC 1


RESULT 7
CG411076
LOCUS       CG411076                663 bp    DNA     linear   GSS 01-NOV-2003
DEFINITION  RM391 Lxx transposon mutant library Leifsonia xyli subsp. xyli
            genomic, genomic survey sequence.
ACCESSION   CG411076
VERSION     CG411076.1  GI:38142787
KEYWORDS    GSS.
SOURCE      Leifsonia xyli subsp. xyli
  ORGANISM  Leifsonia xyli subsp. xyli
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Micrococcineae; Microbacteriaceae; Leifsonia.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Brumbley,S.M., Petrasovits,L.A., Murphy,R.M., Nagel,R.J.,
            Candy,J.M. and Hermann,S.R.
  TITLE     Establishment of a functional genomics platform for Leifsonia xyli
            subsp. xyli
  JOURNAL   Mol. Plant Microbe Interact. 17 (2), 175-183 (2004)
   PUBMED   14964531
COMMENT     Contact: Brumbley SM
            DNPRC
            BSES Ltd
            50 Meiers Rd, Indooroopilly, Brisbane, 4068, Australia
            Tel: 61 7 3331 3333
            Fax: 61 7 3871 0383
            Email: sbrumbley@bses.org.au
            Class: transposon insertion site.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Leifsonia xyli subsp. xyli"
                     /mol_type="genomic DNA"
                     /sub_species="xyli"
                     /db_xref="taxon:59736"
                     /clone_lib="Lxx transposon mutant library"
                     /note="Vector: pUCD623; Lxx transposon mutants were
                     generated by electroporating pUCD623 containing the
                     transposon Tn4431 into Lxx cells grown from 10-15 days in
                     S8 broth containing 0.1% glycine."
ORIGIN

  Query Match             13.7%;  Score 658;  DB 10;  Length 663;
  Best Local Similarity   100.0%;  Pred. No. 1e-172;
  Matches  658;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2118 ATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTA 2177
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTA 60

Qy       2178 TTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTC 2237
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTC 120

Qy       2238 GGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTG 2297
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTG 180

Qy       2298 GCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGC 2357
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGC 240

Qy       2358 ATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGA 2417
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGA 300

Qy       2418 CAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATC 2477
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTCGATCATTGGACCGCTGATC 360

Qy       2478 GTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGC 2537
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGC 420

Qy       2538 GCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACC 2597
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACC 480

Qy       2598 TCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAG 2657
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTCACCACTCCAAGAATTGGAG 540

Qy       2658 CCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCA 2717
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCA 600

Qy       2718 TCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTG 2775
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCGCGTCCGCCATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTG 658


RESULT 8
AJ762384
LOCUS       AJ762384                648 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  AJ762384 G00006 Gerbera hybrid cv. 'Terra Regina' cDNA clone
            G0000600011D12F1, mRNA sequence.
ACCESSION   AJ762384
VERSION     AJ762384.1  GI:62943595
KEYWORDS    EST.
SOURCE      Gerbera hybrid cv. 'Terra Regina'
  ORGANISM  Gerbera hybrid cv. 'Terra Regina'
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Mutisioideae;
            Mutisieae; Gerbera.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Laitinen,R.A., Immanen,J., Auvinen,P., Rudd,S., Alatalo,E.R.,
            Paulin,L., Ainasoja,M., Kotilainen,M., Koskela,S., Teeri,T.H. and
            Elomaa,P.
  TITLE     Analysis of the floral transcriptome uncovers new regulators of
            organ determination and gene families related to flower organ
            differentiation in Gerbera hybrida (Asteraceae)
  JOURNAL   Genome Res. 15 (4), 475-486 (2005)
   PUBMED   15781570
COMMENT     Contact: Alatalo ER
            Institute of Biotechnology
            Viikinkaari 9, P.O. Box 56, 00014 University of Helsinki, FINLAND.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Gerbera hybrid cv. 'Terra Regina'"
                     /mol_type="mRNA"
                     /cultivar="Terra Regina"
                     /db_xref="taxon:226891"
                     /clone="G0000600011D12F1"
```

```
                     /tissue_type="flower stem"
                     /clone_lib="G00006"
ORIGIN

  Query Match             13.5%;  Score 648;  DB 1;  Length 648;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-170;
  Matches  648;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1334 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTG 1393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTG 60

Qy       1394 CTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCAC 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCAC 120

Qy       1454 CGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCG 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCG 180

Qy       1514 GGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGC 240

Qy       1574 GTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCG 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCG 300

Qy       1634 CCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAC 360

Qy       1694 ACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG 420

Qy       1754 TGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT 480

Qy       1814 CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACT 540

Qy       1874 GTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA 600

Qy       1934 CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCG 1981
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCG 648


RESULT 9
AJ762542
LOCUS       AJ762542                 648 bp    mRNA    linear   EST 03-JUN-2005
DEFINITION  AJ762542 G00006 Gerbera hybrid cv. 'Terra Regina' cDNA clone
            G0000600013D05F1, mRNA sequence.
ACCESSION   AJ762542
VERSION     AJ762542.1  GI:62943753
KEYWORDS    EST.
SOURCE      Gerbera hybrid cv. 'Terra Regina'
  ORGANISM  Gerbera hybrid cv. 'Terra Regina'
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; campanulids; Asterales; Asteraceae; Mutisioideae;
            Mutisieae; Gerbera.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Laitinen,R.A., Immanen,J., Auvinen,P., Rudd,S., Alatalo,E.R.,
            Paulin,L., Ainasoja,M., Kotilainen,M., Koskela,S., Teeri,T.H. and
            Elomaa,P.
  TITLE     Analysis of the floral transcriptome uncovers new regulators of
            organ determination and gene families related to flower organ
            differentiation in Gerbera hybrida (Asteraceae)
  JOURNAL   Genome Res. 15 (4), 475-486 (2005)
   PUBMED   15781570
COMMENT     Contact: Alatalo ER
            Institute of Biotechnology
            Viikinkaari 9, P.O. Box 56, 00014 University of Helsinki, FINLAND.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Gerbera hybrid cv. 'Terra Regina'"
                     /mol_type="mRNA"
                     /cultivar="Terra Regina"
                     /db_xref="taxon:226891"
                     /clone="G0000600013D05F1"
                     /tissue_type="flower stem"
                     /clone_lib="G00006"
ORIGIN

  Query Match             13.5%;  Score 648;  DB 1;  Length 648;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-170;
  Matches  648;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1334 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTG 1393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTG 60

Qy       1394 CTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCAC 1453
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCAC 120

Qy       1454 CGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCG 1513
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCG 180

Qy       1514 GGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGC 1573
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGC 240

Qy       1574 GTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCG 1633
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCG 300

Qy       1634 CCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAC 1693
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAC 360

Qy       1694 ACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG 1753
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG 420

Qy       1754 TGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT 1813
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT 480

Qy       1814 CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACT 1873
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACT 540

Qy       1874 GTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA 1933
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA 600

Qy       1934 CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCG 1981
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCG 648


RESULT 10
BF632491
LOCUS       BF632491                 659 bp    mRNA    linear   EST 19-DEC-2000
DEFINITION  NF027F11DT1F1092 Drought Medicago truncatula cDNA clone NF027F11DT
            5', mRNA sequence.
ACCESSION   BF632491
VERSION     BF632491.1  GI:11896649
```

```
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   Torrez-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula drought library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: May GD
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 224 6650
            Fax: 580 224 6692
            Email: gdmay@noble.org
            Insert Length: 659   Std Error: 0.00
            Plate: 027  row: F  column: 11
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF027F11DT"
                     /tissue_type="Plantlets"
                     /dev_stage="Pooled timepoints"
                     /clone_lib="Drought"
                     /note="Vector: Lambda Zap; Contains a mixture of entire
                     plantlets harvested in a series of days-post-watering
                     timepoints. "
ORIGIN

  Query Match             13.4%;  Score 646;  DB 2;  Length 659;
  Best Local Similarity   99.8%;  Pred. No. 2.4e-169;
  Matches  657;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1329 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 1388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTA 60

Qy       1389 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 1448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTC 120

Qy       1449 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 1508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTC 180

Qy       1509 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 1568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTA 240

Qy       1569 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 1628
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGC 300

Qy       1629 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 1688
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCG 360

Qy       1689 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 1748
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCACACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCC 420

Qy       1749 ACA-GGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCG 1807
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACAGGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCG 480

Qy       1808 CCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGG 1867
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGG 540

Qy       1868 GGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGG 1927
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGG 600

Qy       1928 CCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGA 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGA 658


RESULT 11
BQ752029
LOCUS       BQ752029                645 bp    mRNA    linear   EST 18-JUL-2002
DEFINITION  EST632592 DSCT Colletotrichum trifolii cDNA clone pDSCT9-44, mRNA
            sequence.
ACCESSION   BQ752029
VERSION     BQ752029.1  GI:21907434
KEYWORDS    EST.
SOURCE      Colletotrichum trifolii
  ORGANISM  Colletotrichum trifolii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
            mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
            Cheung,F. and Fraser,C.M.
  TITLE     ESTs from mycelia of Colletotrichum trifolii race 1
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            TIGR sequence name: MTSAI44TV More information is available at:
            www.medicago.org
            Seq primer: (gtA AtA CgA CtC ACt AtA ggg C).
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Colletotrichum trifolii"
                     /mol_type="mRNA"
                     /strain="race 1"
                     /db_xref="taxon:5466"
                     /clone="pDSCT9-44"
                     /tissue_type="mycelia"
                     /dev_stage="Young, actively growing mycelia (3 days after
                     inoculation) grown in liquid culture (cutin minimal medium
                     containing 2%glucose)."
                     /lab_host="DH5alpha"
                     /clone_lib="DSCT"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+
                     enriched RNA The cDNA was ligated into Lambda gt11 from
                     Stratagene and packaged using Gigapack packaging extracts.
                     An aliquot of the amplified library was used to transduce
                     E. coli Y1090 and phage DNA was purified from a liquid
                     lysate. The cDNA inserts were gel purified after EcoRI
                     digestion and ligated into pBluescript SK+. Aliquots of
                     the ligation were used to transform E. coli DH5alpha which
                     were plated onto medium with X-gal for selection of
                     recombinants."
ORIGIN

  Query Match             13.4%;  Score 645;  DB 5;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-169;
  Matches  645;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1334 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTG 1393
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 TCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTG 60

Qy       1394 CTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCAC 1453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCAC 120

Qy       1454 CGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCG 1513
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCG 180

Qy       1514 GGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGC 1573
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGC 240

Qy       1574 GTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCG 1633
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCG 300

Qy       1634 CCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAC 1693
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCCAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCAC 360

Qy       1694 ACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG 1753
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGG 420

Qy       1754 TGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT 1813
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTT 480

Qy       1814 CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACT 1873
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACT 540

Qy       1874 GTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA 1933
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAA 600

Qy       1934 CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGA 1978
              ||||||||||||||||||||||||||||||||||||||||||||
Db        601 CCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGA 645


RESULT 12
CG804998/c
LOCUS       CG804998                681 bp    DNA     linear   GSS 10-NOV-2003
DEFINITION  1118056B12.1EL_x1 1118 - RescueMu Grid S Zea mays genomic, genomic
            survey sequence.
ACCESSION   CG804998
VERSION     CG804998.1  GI:38240986
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Very probable ligation site of ends cut by single endonuclease.
            Reverse complemented post-ligation sequence from source sequence.
            Plate: 1118056  row: 40
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1118 - RescueMu Grid S"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid S was grown at San Diego in 2002. DNA was
                     extracted from leaf strips, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
ORIGIN

  Query Match             13.4%;  Score 643;  DB 10;  Length 681;
  Best Local Similarity   99.0%;  Pred. No. 1.7e-168;
  Matches  668;  Conservative    0;  Mismatches    5;  Indels    2;  Gaps    2;

Qy       1706 GATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGG 1765
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        681 GATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGG 622

Qy       1766 CGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAG 1825
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        621 CGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAG 562

Qy       1826 CGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCAT 1885
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 CGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCAT 502

Qy       1886 CTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGG 1945
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 CTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGG 442

Qy       1946 CTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTT 2005
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        441 CTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTT 382

Qy       2006 CAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGAC 2065
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 CAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGAC 322

Qy       2066 TGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGG 2125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 TGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGG 262

Qy       2126 CGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAAT 2185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 CGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAAT 202

Qy       2186 CTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAA 2245
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||| |
Db        201 CTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGGCCCGCCACCAAACGTTTCGGCGAGTA 142

Qy       2246 GCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGC 2305
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        141 GCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGCTACATCTTGCTGGCGTTCGC 82

Qy       2306 GACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGAT 2365
              |||||||||||||||||||| |||||||||||||||||||| | ||||||||||||||||
Db         81 GACGCGAGGCTGGATGGCC-TCCCCATTATGATTCTTCTCGC-TACGGCGGCATCGGGAT 24

Qy       2366 GCCCGCGTTGCAGGC 2380
              ||||||||||||| |
Db         23 GCCCGCGTTGCAGCC 9
```

```
RESULT 13
BJ684207
LOCUS       BJ684207                675 bp    mRNA    linear   EST 25-JAN-2005
DEFINITION  BJ684207 HCEST library Haplochromis chilotes cDNA clone no90g11,
            mRNA sequence.
ACCESSION   BJ684207
VERSION     BJ684207.1  GI:46527328
KEYWORDS    EST.
SOURCE      Haplochromis chilotes
  ORGANISM  Haplochromis chilotes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
            Labroidei; Cichlidae; African cichlids; Pseudocrenilabrinae;
            Haplochromini; Haplochromis.
REFERENCE   1  (bases 1 to 675)
  AUTHORS   Watanabe,M., Kobayashi,N., Shin-i,T., Horiike,T., Tateno,Y.,
            Kohara,Y. and Okada,N.
  TITLE     Extensive analysis of ORF sequences from two different cichlid
            species in Lake Victoria provides molecular evidence for a recent
            radiation event of the Victoria species flock: identity of EST
            sequences between Haplochromis chilotes and Haplochromis sp.
  JOURNAL   Gene 343 (2), 263-269 (2004)
   PUBMED   15588581
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .675
                     /organism="Haplochromis chilotes"
                     /mol_type="mRNA"
                     /db_xref="taxon:257977"
                     /clone="no90g11"
                     /tissue_type="jaw"
                     /dev_stage="varied"
                     /clone_lib="HCEST library"
ORIGIN

  Query Match             13.3%;  Score 641;  DB 3;  Length 675;
  Best Local Similarity   98.4%;  Pred. No. 6.1e-168;
  Matches  658;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy       1336 ATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 1395
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db             8 ATGTTTGACAGCTTAT-ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 66

Qy       1396 AACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCG 1455
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            67 AACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCG 126

Qy       1456 TCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGG 1515
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           127 TCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGG 186

Qy       1516 ATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGT 1575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           187 ATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGT 246

Qy       1576 TGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCC 1635
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           247 TGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCC 306

Qy       1636 CAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACAC 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           307 CAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACAC 366

Qy       1696 CCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTG 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           367 CCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTG 426

Qy       1756 CGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCG 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           427 CGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCG 486

Qy       1816 GGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGT 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           487 GGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGT 546

Qy       1876 TGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACC 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           547 TGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACC 606

Qy       1936 TACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCT 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           607 TACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCTGCAGGCA 666

Qy       1996 TGAGAGCCT 2004
             ||  ||| |
Db           667 TGCAAGCTT 675


RESULT 14
BJ683711
LOCUS       BJ683711                676 bp    mRNA    linear   EST 25-JAN-2005
DEFINITION  BJ683711 HCEST library Haplochromis chilotes cDNA clone no86a05,
            mRNA sequence.
ACCESSION   BJ683711
VERSION     BJ683711.1  GI:46526832
KEYWORDS    EST.
SOURCE      Haplochromis chilotes
  ORGANISM  Haplochromis chilotes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
            Labroidei; Cichlidae; African cichlids; Pseudocrenilabrinae;
            Haplochromini; Haplochromis.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   Watanabe,M., Kobayashi,N., Shin-i,T., Horiike,T., Tateno,Y.,
            Kohara,Y. and Okada,N.
  TITLE     Extensive analysis of ORF sequences from two different cichlid
            species in Lake Victoria provides molecular evidence for a recent
            radiation event of the Victoria species flock: identity of EST
            sequences between Haplochromis chilotes and Haplochromis sp.
  JOURNAL   Gene 343 (2), 263-269 (2004)
   PUBMED   15588581
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Haplochromis chilotes"
                     /mol_type="mRNA"
                     /db_xref="taxon:257977"
                     /clone="no86a05"
                     /tissue_type="jaw"
                     /dev_stage="varied"
                     /clone_lib="HCEST library"
ORIGIN

  Query Match             13.3%;  Score 641;  DB 3;  Length 676;
  Best Local Similarity   98.4%;  Pred. No. 6.1e-168;
  Matches  658;  Conservative    0;  Mismatches   10;  Indels    1;  Gaps    1;

Qy       1336 ATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 1395
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db             8 ATGTTTGACAGCTTAT-ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCT 66
```

```
Qy       1396 AACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCG 1455
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 AACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCG 126

Qy       1456 TCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGG 1515
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 TCACCCTGGATGCTGTAGGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGG 186

Qy       1516 ATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGT 1575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 ATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGT 246

Qy       1576 TGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCC 1635
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 TGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCC 306

Qy       1636 CAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACAC 1695
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAGTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACAC 366

Qy       1696 CCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTG 1755
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 CCGTCCTGTGGATCCTCTACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTG 426

Qy       1756 CGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCG 1815
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 CGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCG 486

Qy       1816 GGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGT 1875
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGT 546

Qy       1876 TGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACC 1935
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACC 606

Qy       1936 TACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCT 1995
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||        |
Db        607 TACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCTGCAGGCA 666

Qy       1996 TGAGAGCCT 2004
             ||  ||| |
Db        667 TGCAAGCTT 675


RESULT 15
BJ683533
LOCUS       BJ683533                683 bp    mRNA    linear   EST 25-JAN-2005
DEFINITION  BJ683533 HCEST library Haplochromis chilotes cDNA clone no82a09,
            mRNA sequence.
ACCESSION   BJ683533
VERSION     BJ683533.1  GI:46526654
KEYWORDS    EST.
SOURCE      Haplochromis chilotes
  ORGANISM  Haplochromis chilotes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
            Labroidei; Cichlidae; African cichlids; Pseudocrenilabrinae;
            Haplochromini; Haplochromis.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Watanabe,M., Kobayashi,N., Shin-i,T., Horiike,T., Tateno,Y.,
            Kohara,Y. and Okada,N.
  TITLE     Extensive analysis of ORF sequences from two different cichlid
            species in Lake Victoria provides molecular evidence for a recent
            radiation event of the Victoria species flock: identity of EST
            sequences between Haplochromis chilotes and Haplochromis sp.
  JOURNAL   Gene 343 (2), 263-269 (2004)
   PUBMED   15588581
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Haplochromis chilotes"
                     /mol_type="mRNA"
                     /db_xref="taxon:257977"
                     /clone="no82a09"
                     /tissue_type="jaw"
                     /dev_stage="varied"
                     /clone_lib="HCEST library"
ORIGIN

  Query Match             13.2%;  Score 637.4;  DB 3;  Length 683;
  Best Local Similarity   98.3%;  Pred. No. 6.3e-167;
  Matches  644;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1353 ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCG 1412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 ATCGATAAGCTTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCG 83

Qy       1413 TGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTA 1472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 TGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTA 143

Qy       1473 GGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGAC 1532
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 GGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGAC 203

Qy       1533 AGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGC 1592
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGC 263

Qy       1593 GCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCG 1652
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 GCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCG 323

Qy       1653 CTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTC 1712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 CTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTC 383

Qy       1713 TACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTAT 1772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 TACGCCGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTAT 443

Qy       1773 ATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGT 1832
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 ATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGT 503

Qy       1833 TTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTG 1892
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 TTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTG 563

Qy       1893 CATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTC 1952
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 CATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGGCTGCTTC 623

Qy       1953 CTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCA 2007
             ||||||||||||||||||||||||||||||||||||      | ||  ||| | ||
Db        624 CTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCTGCAGGCATGCAAGCTTGCA 678


Search completed: January 18, 2006, 11:40:55
Job time : 20380.3 secs
```